Express Mail Label No. EI443365741US

# IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

# UTILITY APPLICATION AND FEE TRANSMITTAL (1.53(b)

ASSISTANT COMMISSIONER FOR PATENTS

	ent Application ton, D.C. 20231				
Sir:					
Transmi	tted herewith for filing is the patent application of				
	med Inventor ication Identifier:    Jens Bukh, 5805 Sonoma Road, Maryland 20817   Roger H. Miller, 15504 White Willow Lane, Rockville, Maryland 20853   Robert H. Purcell, 17517 White Grounds Road, Boyds, Maryland 20841				
For:	NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES				
Enclose	d are:				
	253 page(s) of specification, including <u>162</u> pages of Sequence Listing, <u>1</u> page(s) of Abstract, <u>10</u> page(s) of claims				
[X]_	89 sheets of drawing [ ] formal [ X ] informal				
[ X ] _	2 page(s) of Declaration and Power of Attorney				
	[ ] Unsigned [ ] Newly Executed [X] Copy from prior application				
	[ ] Deletion of inventors including Signed Statement under 37 C.F.R. § 1.63(d)(2)				
[ X ]	Incorporation by Reference: The entire disclosure of the prior application, from which a copy of the combined declaration and power of attorney is supplied herein, is considered as being part of the disclosure of the accompanying application and is incorporated herein by reference.				
[ ]	Microfiche Computer Program (Appendix)				
[ X ]	162 page(s) of Sequence Listing				
	[X] computer readable disk containing Sequence Listing				
	[ X ] Statement under 37 C.F.R. § 1.821(f) that computer and paper copies of the Sequence Listing are the same				
[ ]	Certified copy of Priority Document(s)				
	[ ] English translation documents				

[ X ]	Information Disclosure Statement					
	[ ] Copy of cited references					
[ ]	Preliminary Amendment					
[ X ]	Return receipt postcard (MPEP 503)					
[ X ]	Assignment Papers (assignment cover sheet and assignment documents)					
	[ ] A check in the amount of \$40.00 for recording the Assignment.					
	[ X ] Assignment papers filed in parent application Serial No. <u>08/290,665</u>					
	[ ] Certification of chain of title pursuant to 37 C.F.R. § 3.73(b).					
[ X ]	This is a [ ] continuation [ X ] divisional [ ] continuation-in-part (C-I-P) of prior application serial no.					
	[ $X$ ] Cancel in this application original claims $\underline{21\text{-}31}$ of the parent application before calculating the filling fee. (At least one original independent claim must be retained for filling purposes.)					
	[ ] A Preliminary Amendment is enclosed. (Claims added by this Amendment have been properly numbered consecutively beginning with the number following the highest numbered original claim in the prior application.					
[ X ]	The status of the parent application is as follows:					
	[ ] A Petition For Extension of Time and a Fee therefor has been or is being filed in the parent application to extend the term for action in the parent application until					
	[ ] A copy of the Petition for Extension of Time in the co-pending parent application is attached.					
	[X] No Petition For Extension of Time and Fee therefor are necessary in the co-pending parent application.					
l l	Please abandon the parent application at a time while the parent application is pending or at a time when the petition for extension of time in that application is granted and while this application is pending has been granted a filing date, so as to make this application co-pending.					
	[ ] Transfer the drawing(s) from the parent application to this application.					
[ X ]	Amend the specification by inserting before the first line the sentence:  This is a [ ] continuation [X] divisional [ ] continuation-in-part of co-pending application Seria No. 08/209,665 [illed August 15, 1994 which is a continuation-in-part of application Serial No. 08/086,428 [illed on June 29, 1993 , now U.S. Patent No5,514,539					

Richard W. Bork Registration No. 36,459

	I. CALCU	LATION OF A	PPLICAT	TION FEE				
							Basic Fee	
		Number Filed		Number Extra		Rate	\$ 790.00	
Total								
Claims		55	-20=_	35	х	\$22.00	\$ 770.00	
Indeper		18	- 3=	15	х	\$82.00	\$1230.00	
Multipl	e Dependent Clair	ns [X]y		Additional fee Additional fee	= =	\$270.00 NONE	\$ 270.00	
						Total:	\$3060.00	
A statement claiming small entity status is attached or has been filed in the above-identified parent application and its benefit under 37 C.F.R. § 1.28(a) is hereby claimed. Reduced fees under 37 C.F.R. § 1.9(F) (50% of total) paid herewith \$  X A check in the amount of \$ in payment of the application filing fees is attached.  Charge Fee(s) to Deposit Account No. 13-4500. Order No A DUPLICATE COPY OF THIS SHEET IS ATTACHED.								
[ X ]	The Assistant Commissioner is hereby authorized to charge any additional fees which may be required for filing this application, or credit any overpayment to Deposit Account No. 13-4500 Order No. 2026-4116US2. A DUPLICATE COPY OF THIS SHEET IS ATTACHED.							
					Respec	etfully submitted,		
					MORG	GAN & FINNEG	AN, L.L.P.	
Dated:	May 26, 1998			Ву:		Rickel &	w Bork	

CORRESPONDENCE ADDRESS:

Morgan & Finnegan L.L.P. 345 Park Avenue New York, New York 10154 (212) 758-4800 (212) 751-6849 Facsimile

FORM: UTL-TRAN.NY Rev. 1/16/98

#### IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Bukh, et al. Group Art Unit: To Be Assigned

Serial No. : To Be Assigned Examiner : To Be Assigned

Filed : May 26, 1998

For : NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE ENVELOPE 1

AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN DIAGNOSTIC METHODS AND

VACCINES

Assistant Commissioner of Patents and Trademarks Washington, D.C. 20231

#### TRANSMITTAL REMARKS

Sir:

This paper is submitted to inform the Examiner of certain minor changes that have been made to the copy of the specification of parent application number 08/290,665 submitted herewith.

In particular, the claim to priority on page 1 of the application and the prior and current application data in the Sequence Listing have been updated. In addition, the number "585" at the end of the last line of SEO ID NO:154 in the Sequence Listing has been changed to "573".

These changes are believed to be ministerial in nature and do not therefore result in the introduction of new matter into the specification.

The Commissioner is hereby authorize to charge any fees which may be required for this paper or credit any overpayment to deposit account number 13-4500, order number 2026-4116US2. A DUPLICATE OF THIS SHEET IS ATTACHED.

Early and favorable action by the Examiner is earnestly solicited.

Respectfully submitted,

MORGAN & FINNEGAN, L.L.P.

Dated: May 26, 1998 By: See But Richard W. Bork Registration No. 36,459

# Mailing Address:

MORGAN & FINNEGAN, L.L.P 345 Park Avenue New York, New York 10154 (212) 758-4800 (212) 751-6849 Telecopier

357768\_1

# IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Bukh, et al. Group Art Unit: To Be Assigned

Serial No. : To Be Assigned Examiner : To Be Assigned

Filed : May 26, 1998

For : NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE ENVELOPE 1

AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN DIAGNOSTIC METHODS AND

VACCINES

STATEMENT UNDER 37 C.F.R. §1.821(f) or §1.825(b)						
Assistant Commissioner of Patents and Trademarks Washington, D.C. 20231						
Sir:						
I hereby certify that:						
[X] The paper Sequence Listing and computer readable Sequence Listing submitted herewith are identical						
(37. C.F.R. §1.821(f)).						
[ ] The substitute paper Sequence Listing and substitute computer readable Sequence Listing submittee herewith are identical. No new matter is included (37 C.F.R. §1.825(b)).  Respectfully submitted,						
Dated: May 26, 1998	MORGAN & FINNEGAN, L.L.P.  By: Perkend are food  Richard W. Bork Registration No. 36,459					
Mailing Address:						
MORGAN & FINNEGAN, L.L.P 345 Park Avenue New York, New York 10154 (212) 758-4800 (212) 751-6849 Telecopier						

#### Title of the Invention

NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE ENVELOPE I AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

The present application is a divisional application of pending U.S. Application Serial No. 08/290,665, filed August 15, 1994, which is a continuation-in-part of U.S. Application Serial No. 08/086,428, filed on June 29, 1993, now U.S. Patent No. 5,514,539.

Field Of Invention

The present invention is in the field of hepatitis virology. The invention relates to the complete nucleotide and deduced amino acid sequences of the envelope 1 (E1) and core genes of hepatitis C virus (HCV) isolates from around the world and the grouping of these isolates into fourteen distinct HCV genotypes. More specifically, this invention relates to oligonucleotides, peptides and recombinant proteins derived from the envelope 1 and core gene sequences of these isolates of hepatitis C virus and to diagnostic methods and vaccines which employ these reacents.

## 25 Background Of Invention

Hepatitis C, originally called non-A, non-B hepatitis, was first described in 1975 as a disease serologically distinct from hepatitis A and hepatitis B (Feinstone, S.M. et al. (1975) N. Engl. J. Med. 292:767-770). Although hepatitis C was (and is) the leading type of transfusion-associated hepatitis as well as an important part of community-acquired hepatitis, little progress was made in understanding the disease until the recent identification of hepatitis C virus (HCV) as the causative agent of hepatitis C via the cloning and sequencing of the

372577 1

35

30

5

10

15

HCV genome (Choo, A.L. et al. (1989) Science 288:359-362). The sequence information generated by this study resulted in the characterization of HCV as a small, enveloped, positive-stranded RNA virus and led to the demonstration that HCV is a major cause of both acute and chronic hepatitis worldwide (Weiner, A.J. et al. (1990) Lancet 5 335:1-3). These observations, combined with studies showing that over 50% of acute cases of hepatitis C progress to chronicity with 20% of these resulting in cirrhosis and an undetermined proportion progressing to liver cancer, have led to tremendous efforts by 10 investigators within the hepatitis C field to develop diagnostic assays and vaccines which can detect and prevent hepatitis C infection.

The cloning and sequencing of the HCV genome by Choo et al. (1989) has permitted the development of 15 serologic tests which can detect HCV or antibody to HCV (Kuo, G. et al. (1989) Science 244:362-364). In addition, the work of Choo et al. has also allowed the development of methods for detecting HCV infection via amplification of HCV RNA sequences by reverse transcription and cDNA 20 polymerase chain reaction (RT-PCR) using primers derived from the HCV genomic sequence (Weiner, A.J. et al.). However, although the development of these diagnostic methods has resulted in improved diagnosis of HCV infection, only approximately 60% of cases of hepatitis C 25 are associated with a factor identified as contributing to transmission of HCV (Alter, M.J. et al. (1989) JAMA 262:1201-1205). This observation suggests that effective control of hepatitis C transmission is likely to occur only via universal pediatric vaccination as has been initiated 30 recently for hepatitis B virus. Unfortunately, attempts to date to protect chimpanzees from hepatitis C infection via administration of recombinant vaccines have had only limited success. Moreover, the apparent genetic heterogeneity of HCV, as indicated by the recent assignment 35

For example, one possible obstacle to the development of effective hepatitis C vaccines would arise

5

10

15

20

25

30

35

of all available HCV isolates to one of four genotypes, IIV (Okamoto, H. et al. (1992) J. Gen. Virol; 73:673-679),
presents additional hurdles which must be overcome in order
to develop accurate and effective diagnostic assays and
vaccines.

if the observed genetic heterogeneity of HCV reflects serologic heterogeneity. In such a case, the most genetically diverse strains of HCV may then represent different serotypes of HCV with the result being that infection with one strain may not protect against infection with another. Indeed, the inability of one strain to protect against infection with another strain was recently noted by both Farci et al. (Farci, P. et al. (1992) Science 258:135-140) and Prince et al. (Prince, A.M. et al. (1992) J. Infect. Dis. 165:438-443), each of whom presented evidence that while infection with one strain of HCV does modify the degree of the hepatitis C associated with the reinfection, it does not protect against reinfection with a closely related strain. The genetic heterogeneity among different HCV strains also increases the difficulty encountered in developing RT-PCR assays to detect HCV infection since such heterogeneity often results in falsenegative results because of primer and template mismatch. In addition, currently used serologic tests for detection of HCV or for detection of antibody to HCV are not sufficiently well developed to detect all of the HCV genotypes which might exist in a given blood sample. Finally, in terms of choosing the proper treatment modality to combat hepatitis infection, the inability of presently available serologic assays to distinguish among the various genotypes of HCV represents a significant shortcoming in that recent reports suggest that an HCV-infected patient's response to therapy might be related to the genotype of the infectious virus (Yoshioka, K. et al. (1992) Hepatology

20

25

30

35

16:293-299; Kanai, K. et al. (1992) Lancet 339:1543; Lan, J.Y.N. et al. (1992) Hepatology 16:209A). Indeed, the data presented in the above studies suggest that the closely related genotypes I and II are less responsive to interferon therapy than are the closely related genotypes III and IV. Moreover, preliminary data by Pozzato et al. 5 (Pozzato, G. et al. (1991) Lancet 338:509) suggests that different genotypes may be associated with different types or degrees of clinical disease. Taken together, these studies suggest that before effective vaccines against HCV infection can be developed, and indeed, before more 10 accurate and effective methods for diagnosis and treatment of HCV infection can be produced, one must obtain a greater knowledge about the genetic and serologic diversity of HCV isolates.

In a recent attempt to gain an understanding of the extent of genetic heterogeneity among HCV strains, Bukh et al. carried out a detailed analysis of HCV isolates via the use of PCR technology to amplify different regions of the HCV genome (Bukh, J. et al. (1992a) Proc. Natl. Acad. Sci. 89:187-191). Following PCR amplification, the 5'noncoding (5' NC) portion of the genomes of various HCV isolates were sequenced and it was found that primer pairs designed from conserved regions of the 5' NC region of the HCV genome were more sensitive for detecting the presence of HCV than were primer pairs representing other portions of the genome (Bukh, J. et al. (1992b) Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946). In addition, the authors noted that although many of the HCV isolates examined could be classified into the four genotypes described by Okamoto et al. (1992), other previously undescribed genotypes emerged based on genetic heterogeneity observed in the 5' NC region of the various isolates. One of the most prominent of these newly noted genotypes comprised a group of related viruses that contained the most genetically divergent 5' NC regions of those studied. This group of viruses,

centatively classified as a fifth genotype, are very similar to strains recently described by others (Cha, T.-A et al. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:7144-7148; Chan, S-W. et al. (1992) J. Gen. Virol., 73:1131-1141 and Lee, C-H et al. (1992) J. Clin. Microbio. 30:1602-1604).
In addition, at least four more putative genotypes were identified thereby providing evidence that the genetic heterogeneity of HCV was more extensive than previously appreciated.

However, while the studies of Bukh et al. (1992a and b) provided new and useful information on the genetic heterogeneity of HCV, it is widely appreciated by those skilled in the art that the three structural genes of HCV, core (C), envelope (E1) and envelope 2/nonstructural 1 (E2/NS1) are the most important for the development of serologic diagnostics and vaccines since it is the product of these genes that constitutes the hepatitis C virion. Thus, a determination of the nucleotide sequence of one or all of the structural genes of a variety of HCV isolates would be useful in designing reagents for use in diagnostic assays and vaccines since a demonstration of genetic heterogeneity in a structural gene(s) of HCV isolates might suggest that some of the HCV genotypes represent distinct serotypes of HCV based upon the previously observed relationship between genetic heterogeneity and serologic heterogeneity among another group of single-stranded, positive-sense RNA viruses, the picornaviruses (Ruechert, R.R. "Picornaviridae and their replication", in Fields, B.N. et al., eds. Virology, New York: Raven Press, Ltd. (1990) 507-548).

30

35

10

15

20

25

## Summary of Invention

The present invention relates to cDNAs encoding the complete nucleotide sequence of either the envelope 1 (E1) gene or the core (C) gene of an isolate of human hepatitis C virus (HCV).

^

10

15

20

25

30

35

 $\begin{tabular}{ll} The present invention also relates to the nucleic acid and deduced amino acid sequences of these E1 and core cDNAs. \\ \end{tabular}$ 

It is an object of this invention to provide synthetic nucleic acid sequences capable of directing production of recombinant E1 and core proteins, as well as equivalent natural nucleic acid sequences. Such natural nucleic acid sequences may be isolated from a cDNA or genomic library from which the gene capable of directing synthesis of the E1 or core proteins may be identified and isolated. For purposes of this application, nucleic acid sequence refers to RNA, DNA, cDNA or any synthetic variant thereof which encodes for peptides.

The invention also relates to the method of preparing recombinant E1 and core proteins derived from E1 and core cDNA sequences respectively by cloning the nucleic acid encoding either the recombinant E1 or core protein and inserting the cDNA into an expression vector and expressing the recombinant protein in a host cell.

The invention also relates to isolated and substantially purified recombinant E1 and core proteins and analogs thereof encoded by E1 and core cDNAs respectively.

The invention further relates to the use of recombinant E1 and core proteins, either alone, or in combination with each other, as diagnostic agents and as vaccines.

recombinant production of the core protein of the present invention to contain a second protein on its surface and therefore serve as a carrier in a multivalent vaccine preparation. Further, the present invention relates to the use of the self aggregating core or envelope proteins as a drug delivery system for anti-virals.

The present invention also relates to the

The invention also relates to the use of singlestranded antisense poly- or oligonucleotides derived from El or core cDNAs, or from both El and core cDNAs, to

inhibit expression of hepatitis C E1 and/or core genes.

The invention further relates to multiple computer-generated alignments of the nucleotide and deduced amino acid sequences of the E1 and core cDNAs. These multiple sequence alignments produce consensus sequences which serve to highlight regions of homology and non-homology between sequences found within the same genotype or in different genotypes and hence, these alignments can be used by one skilled in the art to design peptides and oligonucleotides useful as reagents in diagnostic assays and vaccines.

The invention therefore also relates to purified and isolated peptides and analogs thereof derived from E1 and core cDNA sequences.

The invention further relates to the use of these peptides as diagnostic agents and vaccines.

The present invention also encompasses methods of detecting antibodies specific for hepatitis C virus in biological samples. The methods of detecting HCV or antibodies to HCV disclosed in the present invention are useful for diagnosis of infection and disease caused by HCV and for monitoring the progression of such disease. Such methods are also useful for monitoring the efficacy of therapeutic agents during the course of treatment of HCV infection and disease in a mammal.

The invention also provides a kit for the detection of antibodies specific for HCV in a biological sample where said kit contains at least one purified and isolated peptide derived from the E1 or core cDNA sequences. In addition, the invention provides for a kit containing at least one purified and isolated peptide derived from the E1 cDNA sequences and at least one purified and isolated peptide derived from the core cDNA sequences.

The invention further provides isolated and purified genotype-specific oligonucleotides and analogs

30

10

15

20

25

thereof derived from E1 and core cDNA sequences.

The invention also relates to methods for detecting the presence of hepatitis C virus in a mammal, said methods comprising analyzing the RNA of a mammal for the presence of hepatitis C virus. The invention further relates to methods for determining the genotype of hepatitis C virus present in a mammal. This method is useful in determining the proper course of treatment for an HCV-infected patient.

The invention also provides a diagnostic kit for the detection of hepatitis C virus in a biological sample. The kit comprises purified and isolated nucleic acid sequences useful as primers for reverse-transcription polymerase chain reaction (RT-PCR) analysis of RNA for the presence of hepatitis C virus genomic RNA.

The invention further provides a diagnostic kit for the determination of the genotype of a hepatitis C virus present in a mammal. The kit comprises purified and isolated nucleic acid sequences useful as primers for RT-PCR analysis of RNA for the presence of HCV in a biological sample and purified and isolated nucleic acid sequences useful as hybridization probes in determining the genotype of the HCV isolate detected in PCR analysis.

This invention also relates to pharmaceutical compositions useful in prevention or treatment of hepatitis C in a mammal.

## Description of Figures

Figures 1 A-H show computer generated sequence alignments of the nucleotide sequences of 51 HCV E1 cDNAs. The single letter abbreviations used for the nucleotides shown in Figures 1A-H are those standardly used in the art. Figure 1A shows the alignment of SEQ ID NOs:1-8 to produce a consensus sequence for genotype I/la. Figure 1B shows the alignment of SEQ ID NOs:9-25 to produce a consensus sequence for genotype II/lb. Figure 1C shows the alignment

35

5

10

15

20

25

of SEO ID NOs: 26-29 to produce a consensus sequence for genotype III/2a. Figure 1D shows the alignment of SEQ ID NOs:30-33 to produce a consensus sequence for genotype Figure 1E shows the alignment of SEQ ID NOs:35-39 to produce a consensus sequence for genotype V/3a. 1F shows the computer alignment of SEQ ID NOs:42-43 to 5 produce a "consensus" sequence for genotype 4C where the "consensus" sequence given is that of SEQ ID NO:42. Figure 1G shows the alignment of SEQ ID NOs:45-50 to produce a consensus sequence for genotype 5a. The nucleotides shown in capital letters in the consensus sequences of Figures 10 1A-G are those conserved within a genotype while nucleotides shown in lower case letters in the consensus sequences are those variable within a genotype. addition, in Figures 1A-E and 1G, when the lower case letter is shown in a consensus sequence, the lower case 15 letter represents the nucleotide found most frequently in the sequences aligned to produce the consensus sequence. In Figure 1F, the lower case letters shown in the consensus sequence are nucleotides in SEQ ID NO:42 which differ from nucleotides found in the same positions in SEQ ID NO:43. 20 Finally, a hyphen at a nucleotide position in the consensus sequences in Figures 1A-G indicates that two nucleotides were found in equal numbers at that position in the aligned sequences. In the aligned sequences, nucleotides are shown in lower case letters if they differed from the nucleotides 25 of both adjacent isolates. Figure 1H shows the alignment of the consensus sequences of Figures 1A-G with SEQ ID NO:34 (genotype 2c), SEQ ID NO:40 (genotype 4a), SEQ ID NO:41 (genotype 4b), SEQ ID NO:44 (genotype 4d) and SEQ ID NO:51 (genotype 6a) to produce a consensus sequence for all 30 twelve genotypes. This consensus sequence is shown as the bottom line of Figure 1H where the nucleotides shown in capital letters are conserved among all genotypes and a blank space indicates that the nucleotide at that position is not conserved among all genotypes. 35

Figures 2A-H show computer alignments of the deduced amino acid sequences of 51 HCV E1 cDNAs. The single letter abbreviations used for the amino acids shown in Figures 2A-H follow the conventional amino acid shorthand for the twenty naturally occurring amino acids. Figure 2A shows the alignment of SEQ ID NOs:52-59 to produce a consensus sequence for genotype I/la. Figure 2B shows the alignment of SEQ ID NOs:60-76 to produce a consensus sequence for genotype II/1b. Figure 2C shows the alignment of SEQ ID NOs:77-80 to produce a consensus sequence for genotype III/2a. Figure 2D shows the 10 alignment of SEQ ID NOs:81-84 to produce a consensus sequence for genotype IV/2b. Figure 2E shows the alignment of SEO ID NOs:86-90 to produce a consensus sequence for genotype V/3a. Figure 2F shows the computer alignment of SEO ID NOs:93-94 to produce a consensus sequence for 15 genotype 4c. Figure 2G shows the alignment of SEQ ID NOs:96-101 to produce a consensus sequence for genotype 5a. The amino acids shown in capital letters in the consensus sequences of Figures 2A-G are those conserved within a genotype while amino acids shown in lower case letters in 20 the consensus sequences are those variable within a genotype. In addition, in Figures 2A-E and 2G when the lower case letter is shown in a consensus sequence, the letter represents the amino acid found most frequently in the sequences aligned to produce the consensus sequence. 25 In Figure 2F, the lower case letters shown in the consensus sequence are amino acids in SEQ ID NO:93 which differ from amino acids found in the same positions in SEQ ID NO:94. Finally, a hyphen at an amino acid position in the consensus sequences of Figures 2A-G indicates that two 30

consensus sequences of Figures 2A-G indicates that two amino acids were found in equal numbers at that position in the aligned sequences. In the aligned sequences, amino acids are shown in lower case letters if they differed from the amino acids of both adjacent isolates. Figure 2H shows the alignment of the consensus sequences of Figures 2A-G

372577 1

15

20

25

with SEQ ID NO:85 (genotype 2c), SEQ ID NO:91 (genotype 4a), SEQ ID NO:92 (genotype 4b), SEQ ID NO:95 (genotype 4d) and SEQ ID NO:102 (genotype 6a) to produce a consensus sequence for all twelve genotypes. This consensus sequence is shown as the bottom line of Figure 2H where the amino acids shown in capital letters are conserved among all genotypes and a blank space indicates that the amino acid at that position is not conserved among all genotypes.

Figure 3 shows multiple sequence alignment of the deduced amino acid sequence of the El gene of 51 HCV isolates collected worldwide. The consensus sequence of the El protein is shown in boldface (top). In the consensus sequence cysteine residues are highlighted with stars, potential N-linked glycosylation sites are underlined, and invariant amino acids are capitalized, whereas variable amino acids are shown in lower case letters. In the alignment, amino acids are shown in lower case letters if they differed from the amino acid of both adjacent isolates. Amino acid residues shown in bold print in the alignment represent residues which at that position in the amino acid sequence are genotype-specific. acids that were invariant among all HCV isolates are shown as hyphens (-) in the alignment. Amino acid positions correspond to those of the HCV prototype sequence (HCV-1, Choo, L. et al. (1991) Proc. Natl. Acad. Sci. USA 88:2451-

2455) with the first amino acid of the El protein at position 192. The grouping of isolates into 12 genotypes (I/la, II/lb, III/2a, IV/2b, V/3a, 2c, 4a, 4b, 4c, 4d, 5a and 6a) is indicated.

Figure 4 shows a dendrogram of the genetic
relatedness of the twelve genotypes of HCV based on the
percent amino acid identity of the E1 gene of the HCV
genome. The twelve genotypes shown are designated as I/la,
II/lb, III/2a, IV/2b, V/3a, 2c, 4a, 4b, 4c, 4d, 5a and 6a.
The shaded bars represent a range showing the maximum and
minimum homology between the amino acid sequence of any one

Figure 5 shows the distribution of the complete

5

10

15

20

25

30

35

isolate of the genotype indicated and the amino acid sequence of any other isolate.

E1 gene sequence of 74 HCV isolates into the twelve HCV genotypes in the 12 countries studied. For 51 of these HCV isolates, including 8 isolates of genotype I/la, 17 isolates of genotype II/lb and 26 isolates comprising the additional 10 genotypes, the complete E1 gene sequence was determined. In the remaining 23 isolates, all of genotypes I/la and II/lb, the genotype assignment was based on only a partial E1 gene sequence. The partially sequenced isolates did not represent additional genotypes in any of the 12 countries. The number of isolates of a particular genotype is given in each of the 12 countries studied. For ease of viewing, those genotypes designated by two terms (e.g., I/la) are indicated by the latter term (e.g. la). The designations used for each country are: Denmark (DK); Dominican Republic (DR); Germany (D); Hong Kong (HK); India (IND); Sardinia, Italy (S); Peru (P); South Africa (SA); Sweden (SW); Taiwan (T); United States (US); and Zaire (Z). National borders depicted in this figure represent those existing at the time of sampling.

Figures 6A-K show computer generated sequence alignments of the nucleotide sequences of 52 HCV core cDNAs. Single letter abbreviations used for the nucleotides shown in Figures 6A-J are those standardly used in the art. Figure 6A shows the alignment of SEQ ID NOs: 103-108 to produce a consensus sequence for genotype I/Ia. Figure 6B shows the alignment of SEQ ID NOs: 109-124 to produce a consensus sequence for genotype II/lb. Figure 6C shows the alignments of the sequences comprising minor genotypes I/Ia (SEQ ID NOS: 103-108) and II/lb (SEQ ID NOs: 109-124) to produce a consensus sequence for the major genotype, genotype 1. Figure 6D shows the alignment of SEQ ID NOs: 125-128 to produce a consensus sequence for genotype III/2a. Figure 6E shows the alignment of SEQ ID

NOs: 129-133 to produce a consensus sequence for genotype IV/2b. Figure 6F shows the alignment of the sequences of minor genotypes III/2a (SEQ ID NOs: 125-128), IV/2b (SEQ ID NOs: 129-133) and 2c (SEQ ID NO: 134) to produce a consensus sequence for the major genotype, genotype 2. Figure 6G shows the alignment of SEQ ID NOs: 135-138 to 5 produce a consensus sequence for genotype V/3a. Figure 6H shows the computer alignment of the sequences of minor genotypes 4a-4f (SEQ ID NOs: 139-145) to produce a consensus sequence for the major genotype, genotype 4. Figure 6I shows the alignment of SEQ ID NOs: 146-153 to 10 produce a consensus sequence for genotype 5a. The nucleotides shown in capital letters in the consensus sequences in Figure 6A-I are those conserved within the genotype while nucleotides shown in lower case letters in the consensus sequences are those variable within a 15 genotype. In addition, when the lower case letter is shown in the consensus sequence, the lower case letter represents the nucleotide found most frequently in the sequences aligned to produce that consensus sequence. Moreover, a hyphen at a nucleotide position in the consensus sequences 20 in Figures 6A-6I indicates that two nucleotides were found in equal numbers at that position in the sequences aligned to produce the consensus sequence. Finally, nucleotides are shown in lower case letters in the sequences aligned to produce each consensus sequence shown in Figures 6A-6I, if 25 they differed from the nucleotides of both adjacent isolates. Figure 6J shows the alignment of the consensus sequences of major genotypes 1 (Figure 6C), 2 (Figure 6F), 3 (Figure 6G), 4 (Figure 6H), 5 (Figure 6I) and 6 (SEQ ID NO: 154) to produce a consensus sequence for all genotypes 30 and Figure 6K shows the alignment of consensus sequences of Figures 6A, 6B, 6D, 6E, 6G and 6I with SEQ ID NO:134 (genotype 2c), SEQ ID NO:139 (genotype 4a), SEQ ID NO:141 (genotype 4b), SEQ ID NO:143 (genotype 4c), SEQ ID NO:145 (genotype 4d), SEQ ID NO:142 (genotype 4e), SEQ ID NO:140 35

20

25

30

35

(genotype 4f) and SEQ ID NO:154 (genotype 6a) to produce a consensus sequence for all fourteen genotypes. The nucleotides shown in capital letters in the consensus sequences of Figures 6J and 6K are conserved among all genotypes and the nucleotide shown in lower case letter 5 represent the nucleotides found most frequently in the sequences aligned to produce this consensus sequence. In addition, the presence of a hyphen at a nucleotide position in all fourteen sequences aligned in Figure 6K indicates that the nucleotide found at that position in the aligned sequences is the same as nucleotide shown at the corresponding position in the consensus sequences of Figure 6K.

Figures 7A-7J show computer alignments of the

deduced amino acid sequences of the 52 HCV core cDNAs. single letter abbreviations used for the amino acids shown in Figures 7A-7J follow the conventional amino acid short hand for the twenty natural occurring amino acids. Figure 7A shows the alignment of SEQ ID NOs: 155-160 to produce a consensus sequence for genotype I/la. Figure 7B shows the alignment of SEQ ID NOs: 161-176 to produce a consensus sequence for genotype II/1b. Figure 7C shows the alignment of the sequences comprising minor genotypes I/a (SEQ ID NOS: 155-160) and II/1b (SEQ ID NOS: 161-176) to produce a consensus sequence for the major genotype, genotype 1. Figure 7D shows the alignment of SEQ ID NOs: 177-180 to produce a consensus sequence for genotype III/2a. Figure 7E shows the alignment of SEQ ID NOs: 181-185 to produce a consensus sequence for genotype IV/2b. Figure 7F shows the alignment of the sequences of minor genotypes III/2a (SEQ TD NOS: 177-180), IV/2b (SEQ ID NOS: 181-185) and 2c (SEQ ID NO: 186) to produce a consensus sequence for the major genotype, genotype 2. Figure 7G shows the alignment of SEQ ID NOs: 187-190 to produce a consensus sequence for genotype V/3a. Figure 7H shows the computer alignment of the sequences of minor genotypes 4a-4f (SEQ ID NOs: 191-

197) to produce a consensus sequence for the major genotype, genotype 4. Figure 7I shows the alignment of SEO ID NOs: 198-205 to produce a consensus sequence for genotype 5a. The amino acids shown in capital letters in the consensus sequences of Figures 7A-7I are those conserved within the genotype while amino acids shown in 5 lower case letters in the consensus sequences are those variable within the genotype. In addition, when a lower case letter is found in the consensus sequences shown in Figures 7A-7I, the letter represents the amino acid found most frequently in the sequences aligned to produce that 10 consensus sequence. Moreover, a hyphen in an amino acid position in the consensus sequences of Figures 7A-7I indicates that two amino acids were found in equal numbers at that position in the sequences aligned to produce that consensus sequence. Finally, amino acids are shown in 15 lower case letters in the sequences aligned to produce the consensus sequences shown in Figures 7A-7I if these amino acids differed from the amino acids of both adjacent isolates. Figure 7J shows the alignment of the consensus sequences of major genotypes 1 (Figure 7C), 2 (Figure 7F), 20 3 (Figure 7G), 4 (Figure 7H), 5 (Figure 7I) and 6 (SEQ ID NO: 154) to produce a consensus sequence for all genotypes and Figure 7K shows the alignment of the consensus sequences of Figures 7A, 7B, 7D, 7E, 7G and 7I with SEQ ID NO:186 (genotype 2c), SEQ ID NO:191 (genotype 4a), SEQ ID 25 NO:193 (genotype 4b), SEQ ID NO:195 (genotype 4c), SEQ ID NO:197 (genotype 4d), SEQ ID NO:194 (genotype 4e), SEQ ID NO:192 (genotype 4f) and SEQ ID NO:206 (genotype 6a) to produce a consensus sequence for all fourteen genotypes. The amino acids shown in capital letters in the consensus 30 sequences shown in Figures 7J and 7K are conserved among all genotypes while the amino acids shown in lower case letters represent amino acids found most frequently in the sequences aligned to produce this consensus sequence. addition, the presence of a hyphen at an amino acid 35

position in all fourteen sequences aligned in Figure 7K indicates that the amino acid found at that position in the aligned sequences is the same as the amino acid shown at the corresponding position in the consensus sequence of Figure 7K.

Figure 8 shows phylogenetic trees illustrating the calculated evolutionary relationships of the different HCV isolates based upon the C gene sequence of 52 HCV isolates and the E1 gene sequence of 51 HCV isolates, respectively. The phylogenetic trees were constructed by the unweighted pair-group method with arithmetic mean (Nei, M. (1987) Molecular Evolutionary Genetics (Columbia University Press, New York, N.Y.), pp 287-326) using the computer software package "Gene Works" from IntelliGenetics. The lengths of the horizontal lines connecting the sequences, given in absolute values from 0 to 1, are proportional to the estimated genetic distances between the sequences. Genotype designations of HCV isolates are indicated. In 45 HCV isolates, both the C and the E1 gene sequences were determined.

т

## Detailed Description Of Invention

The present invention relates to cDNAs encoding the complete nucleotide sequence of the envelope 1 (E1) and core genes of isolates of human hepatitis C virus (HCV). The E1 cDNAs of the present invention were obtained as follows. Viral RNA was extracted from serum collected from humans infected with hepatitis C virus and the viral RNA was then reverse transcribed and amplified by polymerase chain reaction using primers deduced from the sequence of the HCV strain H-77 (Ogata, N. et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396). The amplified cDNA was then isolated by gel electrophoresis and sequenced.

The present invention further relates to the nucleotide sequences of the cDNAs encoding the E1 gene of 51 HCV isolates. These nucleotide sequences are shown in

35

5

10

15

20

25

10

15

20

25

30

35

o the sequence listing as SEQ ID NO:1 through SEQ ID NO:51. The abbreviations used for the nucleotides are those standardly used in the art.

The deduced amino acid sequence of each of SEQ ID NO:1 through SEQ ID NO:51 are presented in the sequence listing as SEQ ID NO:52 through SEQ ID NO:102 where the amino acid sequence in SEQ ID NO:52 is deduced from the nucleotide sequence shown in SEQ ID NO:1, the amino acid sequence shown in SEQ ID NO:1, the amino acid sequence shown in SEQ ID NO:53 is deduced from the nucleotide sequence shown in SEQ ID NO:2 and so on. The deduced amino acid sequence of each of SEQ ID Nos:52-102 starts at nucleotide 1 of the corresponding nucleic acid sequence shown in SEQ ID NOs:1-51 and extends 575 nucleotides to a total length of 576 nucleotides.

The three letter abbreviations used in SEQ ID Nos:52-102 follow the conventional amino acid shorthand for the twenty naturally occurring amino acids.

The present invention also relates to the nucleotide sequences of the cDNAs encoding the core gene of 52 HCV isolates. These nucleotide sequences are shown in the sequence listing as SEQ ID NO:103 through SEQ ID NO:154.

The core cDNAs of the present invention were obtained as follows. Viral RNA was extracted from serum and reversed transcribed as described above for cloning of the El cDNAs. The core cDNAs of the present invention were then amplified by polymerase chain reaction using primers deduced from previously determined sequences that flank the core gene (Bukh et al. (1992)) <a href="Proc. Natl. Acad. Sci.U.S.A.">Proc. Natl. Acad. Sci.U.S.A.</a>, 89: 4942-4946; Bukh et al. (1993) <a href="Proc. Natl. Acad. Sci.U.S.A.">Proc. Natl. Acad. Sci.U.S.A.</a>, 90: 8234-8238).

The deduced amino acid sequence of each of SEQ ID NO:103 through SEQ ID NO:154 are presented in the sequence listing as SEQ ID NO:155 through SEQ ID NO:206 where the amino acid sequence in SEQ ID NO:155 is deduced from the nucleotide sequence shown in SEQ ID NO:103, the amino acid

15

20

25

30

35

sequence shown in SEQ ID NO:156 is deduced from the nucleotide sequence shown in SEQ ID NO:104 and so on. The deduced amino acid sequence of each of SEQ ID NOs: 155-206 starts at nucleotide 1 of the corresponding nucleotide sequence shown in SEQ ID NOs:103-154 and extends 572
nucleotides to a total length of 573 nucleotides.

Preferably, the El and core proteins and peptides of the present invention are substantially homologous to, and most preferably biologically equivalent to, native HCV El and core proteins and peptides. By "biologically equivalent" as used throughout the specification and claims, it is meant that the compositions are immunogenically equivalent to the native E1 and core proteins and peptides. The E1 and core proteins and peptides of the present invention may also stimulate the production of protective antibodies upon injection into a mammal that would serve to protect the mammal upon challenge with HCV. By "substantially homologous" as used throughout the ensuing specification and claims to describe El and core proteins and peptides, it is meant a degree of homology in the amino acid sequence of the E1 and core proteins and peptides to the native E1 and core proteins and peptides respectively. Preferably the degree of homology is in excess of 90, preferably in excess of 95, with a particularly preferred group of proteins being in excess of 99 homologous with the native El or core proteins and peptides.

Variations are contemplated in the cDNA sequences shown in SEQ ID NO:1 through SEQ ID NO:51 and in SEQ ID NO:103 through SEQ ID NO:154 which will result in a nucleic acid sequence that is capable of directing production of analogs of the corresponding protein shown in SEQ ID NO:52 through SEQ ID NO:102 and in SEQ ID NO:155 through SEQ ID NO:0206. It should be noted that the cDNA sequences set forth above represent a preferred embodiment of the present invention. Due to the degeneracy of the genetic code, it

is to be understood that numerous choices of nucleotides may be made that will lead to a DNA sequence capable of directing production of the instant protein or its analogs. As such, DNA sequences which are functionally equivalent to the sequence set forth above or which are functionally equivalent to sequences that would direct production of analogs of the E1 and core proteins produced pursuant to the amino acid sequences set forth above, are intended to be encompassed within the present invention.

The term analog as used throughout the specification or claims to describe the E1 and core proteins and peptides of the present invention, includes any protein or peptide having an amino acid residue sequence substantially identical to a sequence specifically shown herein in which one or more residues have been conservatively substituted with a biologically equivalent residue. Examples of conservative substitutions include the substitution of one polar (hydrophobic) residue such as isoleucine, valine, leucine or methionine for another, the substitution of one polar (hydrophilic) residue for another such as between arginine and lysine, between glutamine and asparagine, between glycine and serine, the substitution of one basic residue such as lysine, arginine or histidine for another, or the substitution of one acidic residue, such as aspartic acid or glutamic acid for another.

25

10

15

20

The phrase "conservative substitution" also includes the use of a chemically derivatized residue in place of a non-derivatized residue provided that the resulting protein or peptide is biologically equivalent to the native E1 or core protein or peptide.

30

"Chemical derivative" refers to an El or core protein or peptide having one or more residues chemically derivatized by reaction of a functional side group.

Examples of such derivatized molecules, include but are not limited to, those molecules in which free amino groups have been derivatized to form amine hydrochlorides, p-toluene

sulfonyl groups, carbobenzoxy groups, t-butyloxycarbonyl groups, chloracetyl groups or formyl groups. Free carboxyl groups may be derivatized to form salts, methyl and ethyl esters or other types of esters or hydrazides. Free hydroxyl groups may be derivatized to form O-acyl or Oalkyl derivatives. The imidazole nitrogen of histidine may 5 be derivatized to form N-imbenzylhistidine. Also included as chemical derivatives are those proteins or peptides which contain one or more naturally-occurring amino acid derivatives of the twenty standard amino acids. For examples: 4-hydroxyproline may be substituted for proline; 10 5-hydroxylysine may be substituted for lysine; 3methylhistidine may be substituted for histidine; homoserine may be substituted for serine; and ornithine may be substituted for lysine. The E1 and core proteins and peptide of the present invention also includes any protein 15 or peptide having one or more additions and/or deletions of residues relative to the sequence of a peptide whose sequence is shown herein, so long as the peptide is biologically equivalent to the native El or core protein or peptide. 20

The present invention also includes a recombinant DNA method for the manufacture of HCV E1 and core proteins. In this method, natural or synthetic nucleic acid sequences may be used to direct the production of E1 and core proteins.

In one embodiment of the invention, the method comprises:

- preparation of a nucleic acid sequence capable of directing a host organism to produce HCV E1 or core protein;
- (b) cloning the nucleic acid sequence into a vector capable of being transferred into and replicated in a host organism, such vector containing operational elements for the nucleic acid sequence;
- (c) transferring the vector containing the

35

25

nucleic acid and operational elements into a host organism capable of expressing the protein;

- (d) culturing the host organism under conditions appropriate for amplification of the vector and expression of the protein; and
  - (e) harvesting the protein.

In another embodiment of the invention, the method for the recombinant DNA synthesis of an HCV E1 protein encoded by any one of the nucleic acid sequences shown in SEQ ID NOs:1-51 comprises:

(a) culturing a transformed or transfected host organism containing a nucleic acid sequence capable of directing the host organism to produce a protein, under conditions such that the protein is produced, said protein exhibiting substantial homology to a native E1 protein isolated from HCV having the amino acid sequence according to any one of the amino acid sequences shown in SEQ ID NOs:52-102 or combinations thereof.

In one embodiment, the RNA sequence of an HCV isolate was isolated and converted to cDNA as follows. Viral RNA is extracted from a biological sample collected from human subjects infected with hepatitis C and the viral RNA is then reverse transcribed and amplified by polymerase chain reaction using primers deduced from the sequence of HCV strain H-77 (Ogata et al. (1991)). Preferred primer sequences are shown as SEQ ID NOs:207-212 in the sequence listing. Once amplified, the PCR fragments are isolated by gel electrophoresis and sequenced.

In an alternative embodiment, the above method may be utilized for the recombinant DNA synthesis of an HCV core protein encoded by any one of the nucleic acid sequences shown in SEQ ID NOS: 103-154, where the protein produced by this method exhibits substantial homology to a native core protein isolated from HCV having amino acid sequence according to any one of the amino acid sequences shown in SEQ ID NOS: 155-206 or combinations thereof.

372577 1

30

5

10

15

20

25

15

20

25

30

35

The vectors contemplated for use in the present invention include any vectors into which a nucleic acid sequence as described above can be inserted, along with any preferred or required operational elements, and which vector can then be subsequently transferred into a host organism and replicated in such organisms. Preferred vectors are those whose restriction sites have been well documented and which contain the operational elements preferred or required for transcription of the nucleic acid sequence.

The "operational elements" as discussed herein include at least one promoter, at least one operator, at least one leader sequence, at least one terminator codon, and any other DNA sequences necessary or preferred for appropriate transcription and subsequent translation of the vector nucleic acid. In particular, it is contemplated that such vectors will contain at least one origin of replication recognized by the host organism along with at least one selectable marker and at least one promoter sequence capable of initiating transcription of the nucleic acid sequence.

In construction of the recombinant expression vectors of the present invention, it should additionally be noted that multiple copies of the nucleic acid sequence of interest (either El or core) and its attendant operational elements may be inserted into each vector. In such an embodiment, the host organism would produce greater amounts per vector of the desired El or core protein. The number of multiple copies of the nucleic acid sequence which may be inserted into the vector is limited only by the ability of the resultant vector due to its size, to be transferred into and replicated and transcribed in an appropriate host microorganism.

Of course, those skilled in the art would readily understand that copies of both core and El nucleic acid sequence may be inserted into single vector such that a

15

20

25

host organism transformed or transfected with said vector would produce both the desired E1 and core proteins. For example, a polysistronic vector in which multiple different E1 and/or core proteins may be expressed from a single vector is created by placing expression of each protein under control of an internal ribosomal entry site (IRES) (Molla, A. et al. Nature, 356:255-257 (1992); Gong, S.K. et al. J. of Virol., 263:1651-1660 (1989)).

In another embodiment, restriction digest fragments containing a coding sequence for E1 or core proteins can be inserted into a suitable expression vector that functions in prokaryotic or eukaryotic cells. By suitable is meant that the vector is capable of carrying and expressing a complete nucleic acid sequence coding for an E1 or core protein. Preferred expression vectors are those that function in a eukaryotic cell. Examples of such vectors include but are not limited to vaccinia virus vectors, adenovirus or herpes viruses. A preferred vector is the baculovirus transfer vector, pBlueBac.

In yet another embodiment, the selected recombinant expression vector may then be transfected into a suitable eukaryotic cell system for purposes of expressing the recombinant protein. Such eukaryotic cell systems include but are not limited to cell lines such as HeLa, MRC-5 or CV-1. A preferred eukaryotic cell system is SF9 insect cells.

The expressed recombinant protein may be detected by methods known in the art including, but not limited to, Coomassie blue staining and Western blotting.

The present invention also relates to

substantially purified and isolated recombinant E1 and core proteins. In one embodiment, the recombinant protein expressed by the SF9 cells can be obtained as a crude lysate or it can be purified by standard protein purification procedures known in the art which may include differential precipitation, molecular sieve chromatography,

15

20

25

30

35

o ion-exchange chromatography, isoelectric focusing, gel electrophoresis and affinity and immunoaffinity chromatography. The recombinant protein may be purified by passage through a column containing a resin which has bound thereto antibodies specific for the open reading frame
5 (ORF) protein.

The present invention further relates to the use of recombinant E1 and core proteins as diagnostic agents and vaccines. In one embodiment, the expressed recombinant proteins of this invention can be used in immunoassays for diagnosing or prognosing hepatitis C in a mammal. For the purposes of the present invention, "mammal" as used throughout the specification and claims, includes, but is not limited to humans, chimpanzees, other primates and the like. In a preferred embodiment, the immunoassay is useful in diagnosing hepatitis C infection in humans.

Immunoassays of the present invention may be

those commonly used by those skilled in the art including, but not limited to, radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme immunoassay, chemiluminescent assay, immunohistochemical assay, immunoprecipitation and the like. Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Edition, Rose and Bigazzi, eds., John Wiley and Sons, 1980 and Campbell et al., Methods of Immunology, W.A. Benjamin, Inc., 1964, both of which are incorporated herein by reference. Such assays may be a direct, indirect, competitive, or noncompetitive immunoassay as described in the art (Oellerich, M. 1984. J. Clin. Chem. Clin. BioChem 22:895-904) Biological samples appropriate for such detection assays include, but are not limited to serum, liver, saliva, lymphocytes or other

 $\label{eq:continuous} In a preferred embodiment, test serum is reacted with a solid phase reagent having surface-bound recombinant HCV E1 and/or core protein(s) as antigen(s). The solid$ 

372577 1

mononuclear cells.

20

2.5

30

35

surface reagent can be prepared by known techniques for attaching protein to solid support material. These attachment methods include non-specific adsorption of the protein to the support or covalent attachment of the protein to a reactive group on the support. After reaction of the antigen with anti-HCV antibody, unbound serum components are removed by washing and the antigen-antibody complex is reacted with a secondary antibody such as labelled anti-human antibody. The label may be an enzyme which is detected by incubating the solid support in the presence of a suitable fluorimetric or calorimetric reagent. Other detectable labels may also be used, such as radiolabels or colloidal gold, and the like.

The HCV E1 and/or core proteins and analogs thereof may be prepared in the form of a kit, alone, or in combinations with other reagents such as secondary antibodies, for use in immunoassays.

In yet another embodiment the recombinant E1 and core proteins or analogs thereof can be used as a vaccine to protect mammals against challenge with hepatitis C. The vaccine, which acts as an immunogen, may be a cell, cell lysate from cells transfected with a recombinant expression vector or a culture supernatant containing the expressed protein. Alternatively, the immunogen is a partially or substantially purified recombinant protein. In yet another embodiment, the immunogen may be a fusion protein comprising core protein and a second, non-core protein joined together such that the core portion of the fusion protein will aggregate and "trap" the second protein on the surface of the particle produced by aggregation of the core protein. (Molecular Biology of the Hepatitis B Virus", McLachlan, A. (1991) CRC Press, Boca Raton, Fla.). Alternatively, the core protein could be mixed with the second protein in vitro to produce particles in which all or part of the second protein was exposed on the surface of the particle. Such particles would then serve as a carrier

10

15

20

30

35

in a multi-valent vaccine preparation. Second proteins or parts thereof which could be mixed with or fused to the core protein include, but are not limited to, HCV E1 and hepatitis B surface antigen.

While it is possible for the immunogen to be administered in a pure or substantially pure form, it is preferable to present it as a pharmaceutical composition, formulation or preparation.

The formulations of the present invention, both for veterinary and for human use, comprise an immunogen as described above, together with one or more pharmaceutically acceptable carriers and optionally other therapeutic ingredients. The carrier(s) must be "acceptable" in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. The formulations may conveniently be presented in unit dosage form and may be prepared by any method well-known in the pharmaceutical art.

the pharmaceutical art.

All methods include the step of bringing into association the active ingredient with the carrier which constitutes one or more accessory ingredients. In general,

the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product into the desired

25 formulation.

Formulations suitable for intravenous intramuscular, subcutaneous, or intraperitoneal administration conveniently comprise sterile aqueous solutions of the active ingredient with solutions which are preferably isotonic with the blood of the recipient. Such formulations may be conveniently prepared by dissolving the solid active ingredient in water containing physiologically compatible substances such as sodium chloride (e.g. 0.1-2.0m), glycine, and the like, and having a buffered pH compatible with physiological conditions to produce an

10

15

20

35

aqueous solution, and rendering said solution sterile. These may be present in unit or multi-dose containers, for example, sealed ampules or vials.

The formulations of the present invention may incorporate a stabilizer. Illustrative stabilizers are preferably incorporated in an amount of 0.10-10,000 parts by weight per part by weight of immunogens. If two or more stabilizers are to be used, their total amount is preferably within the range specified above. These stabilizers are used in aqueous solutions at the appropriate concentration and pH. The specific osmotic pressure of such aqueous solutions is generally in the range of 0.1-3.0 osmoles, preferably in the range of 0.8-1.2. The pH of the aqueous solution is adjusted to be within the range of 5.0-9.0, preferably within the range of 6-8. In formulating the immunogen of the present invention, an anti-adsorption agent may be used.

Additional pharmaceutical methods may be employed to control the duration of action. Controlled release preparations may be achieved through the use of polymer to complex or adsorb the proteins or their derivatives. The controlled delivery may be exercised by selecting appropriate macromolecules (for example polyester, polyamino acids, polyvinyl pyrrolidone, ethylenevinylacetate, methylcellulose,

carboxymethylcellulose, or protamine sulfate) and the concentration of macromolecules as well as the methods of incorporation in order to control release. Another possible method to control the duration of action by controlled-release preparations is to incorporate the proteins, protein analogs or their functional derivatives, into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is

possible to entrap these materials in microcapsules

15

20

25

prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly (methylmethacylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, 5 albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions.

When oral preparations are desired, the compositions may be combined with typical carriers, such as lactose, sucrose, starch, talc, magnesium stearate, crystalline cellulose, methyl cellulose, carboxymethyl cellulose, glycerin, sodium alginate or gum arabic among others.

The E1 and core proteins of the present invention may also be used as a delivery system for anti-virals to prevent or attenuate HCV infection in a mammal by utilizing the property of both proteins to self-aggregate in vitro to "trap" the antiviral within the particles produced via aggregation of the core and El proteins. Examples of antivirals which could be delivered by such a system include, but are not limited to antisense DNA or RNAs.

Vaccination can be conducted by conventional methods. For example, the immunogen or immunogens (e.g. the E1 protein may be administered alone or in combination with the E1 proteins derived from other isolates of HCV) can be used in a suitable diluent such as saline or water, or complete or incomplete adjuvants. Further, the immunogen(s) may or may not be bound to a carrier to make the protein(s) immunogenic. Examples of such carrier molecules include but are not limited to bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), tetanus

30 toxoid, and the like. The immunogen(s) can be administered by any route appropriate for antibody production such as intravenous, intraperitoneal, intramuscular, subcutaneous, The immunogen(s) may be administered once or and the like.

at periodic intervals until a significant titer of anti-HCV 35

10

15

20

25

30

35

antibody is produced. The antibody may be detected in the serum using an immunoassay.

In yet another embodiment, the immunogen may be nucleic acid sequence capable of directing host organism synthesis of E1 and/or core protein(s). Such nucleic acid sequence may be inserted into a suitable expression vector by methods known to those skilled in the art. Expression vectors suitable for producing high efficiency gene transfer in vivo include retroviral, adenoviral and vaccinia viral vectors. Operational elements of such expression vectors are disclosed previously in the present specification and are known to one skilled in the art. Such expression vectors can be administered intravenously, intramuscularly, subcutaneously, intraperitoneally or orally.

In an alternative embodiment, direct gene transfer may be accomplished via intramuscular injection of, for example, plasmid-based eukaryotic expression vectors containing a nucleic acid sequence capable of directing host organism synthesis of E1 and/or core protein(s). Such an approach has previously been utilized to produce the hepatitis B surface antigen in vivo and resulted in an antibody response to the surface antigen (Davis, H.L. et al. (1993) Human molecular Genetics, 2:1847-1851; see also Davis et al. (1993) Human Gene Therapy, 4:151-159 and 733-740).

Doses of E1 and/or core protein(s)-encoding nucleic acid sequence effective to elicit a protective antibody response against HCV infection range from about 1 to about 500  $\mu$ g. A more preferred range being about 1 to about 500  $\mu$ g.

The El and/or core proteins and expression vectors containing a nucleic acid sequence capable of directing host organism synthesis of El and/or core protein(s) may be supplied in the form of a kit, alone, or in the form of a pharmaceutical composition as described

above.

5

10

15

20

25

30

35

The administration of the immunogen(s) of the present invention may be for either a prophylactic or therapeutic purpose. When provided prophylactically, the immunogen(s) is provided in advance of any exposure to HCV or in advance of any symptom of any symptoms due to HCV infection. The prophylactic administration of the immunogen serves to prevent or attenuate any subsequent infection of HCV in a mammal. When provided therapeutically, the immunogen(s) is provided at (or shortly after) the onset of the infection or at the onset of any symptom of infection or disease caused by HCV. The therapeutic administration of the immunogen(s) serves to attenuate the infection or disease.

In addition to use as a vaccine, the compositions can be used to prepare antibodies to HCV E1 and core proteins. The antibodies can be used directly as antiviral agents or they may be used in immunoassays disclosed herein to detect HCV El and core proteins present in patient sera.. To prepare antibodies, a host animal is immunized using the El and/or core proteins native to the virus particle bound to a carrier as described above for vaccines. The host serum or plasma is collected following an appropriate time interval to provide a composition comprising antibodies reactive with the El or core protein of the virus particle. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE Sephadex, or other techniques known to those skilled in the art. The antibodies are substantially free of many of the adverse side effects which may be associated with other anti-viral agents such as drugs.

The antibody compositions can be made even more compatible with the host system by minimizing potential adverse immune system responses. This is accomplished by removing all or a portion of the Fc portion of a foreign

25

30

35

species antibody or using an antibody of the same species as the host animal, for example, the use of antibodies from human/human hybridomas. Humanized antibodies (i.e., nonimmunogenic in a human) may be produced, for example, by replacing an immunogenic portion of an antibody with a 5 corresponding, but nonimmunogenic portion (i.e., chimeric antibodies). Such chimeric antibodies may contain the reactive or antigen-binding portion of an antibody from one species and the Fc portion of an antibody (nonimmunogenic) from a different species. Examples of chimeric antibodies, include but are not limited to, non-human mammal-human 10 chimeras, rodent-human chimeras, murine-human and rat-human chimeras (Robinson et al., International Patent Application 184,187; Taniguchi M., European Patent Application 171,496; Morrison et al., European Patent Application 173,494; Neuberger et al., PCT Application WO 86/01533; Cabilly et 15 al., 1987 Proc. Natl. Acad. Sci. USA 84:3439; Nishimura et al., 1987 Canc. Res. 47:999; Wood et al., 1985 Nature 314:446; Shaw et al., 1988 J. Natl. Cancer Inst. 80:15553,

General reviews of "humanized" chimeric antibodies are provided by Morrison S., 1985 Science 229:1202 and by Oi et al., 1986 BioTechniques 4:214.

all incorporated herein by reference).

Suitable "humanized" antibodies can be alternatively produced by CDR or CEA substitution (Jones et al., 1986 Nature 321:552; Verhoeyan et al., 1988 Science 239:1534; Biedleret al. 1988 J. Immunol. 141:4053, all incorporated herein by reference).

The antibodies or antigen binding fragments may also be produced by genetic engineering. The technology for expression of both heavy and light cain genes in E. coli is the subject of the PCT patent applications; publication number WO 901443, WO901443, and WO 9014424 and in Huse et al., 1989 Science 246:1275-1281.

The antibodies can also be used as a means of enhancing the immune response. The antibodies can be

372577\_1

20

25

30

35

administered in amount similar to those used for other therapeutic administrations of antibody. For example, normal immune globulin is administered at 0.02-0.1 ml/lb body weight during the early incubation period of other viral diseases such as rabies, measles, and hepatitis B to interfere with viral entry into cells. Thus, antibodies reactive with the HCV El and/or core proteins can be passively administered alone or in conjunction with another anti-viral agent to a host infected with an HCV to enhance the immune response and/or the effectiveness of an antiviral drug.

Alternatively, anti-HCV E1 antibodies and anti-HCV core antibodies can be induced by administered anti-idiotype antibodies as immunogens. Conveniently, a purified anti-HCV E1 or anti-HCV core antibody preparation prepared as described above is used to induce anti-idiotype antibody in a host animal, the composition is administered to the host animal in a suitable diluent. Following administration, usually repeated administration, the host produces anti-idiotype antibody. To eliminate an immunogenic response to the Fc region, antibodies produced

immunogenic response to the Fc region, antibodies produced by the same species as the host animal can be used or the Fc region of the administered antibodies can be removed. Following induction of anti-idiotype antibody in the host animal, serum or plasma is removed to provide an antibody composition. The composition can be purified as described above for anti-HCV El and anti-HCV core antibodies, or by affinity chromatography using anti-HCV El or anti-HCV core antibodies bound to the affinity matrix. The anti-idiotype antibodies produced are similar in conformation to the authentic HCV El or core protein and may be used to prepare an HCV vaccine rather than using an HCV El or core protein.

When used as a means of inducing anti-HCV virus antibodies in an animal, the manner of injecting the antibody is the same as for vaccination purposes, namely intramuscularly, intraperitoneally, subcutaneously or the

372577\_1

like in an effective concentration in a physiologically suitable diluent with or without adjuvant. One or more booster injections may be desirable.

The HCV E1 and core proteins of the invention are also intended for use in producing antiserum designed for pre- or post-exposure prophylaxis. Here an El or core protein, or mixture of E1 and/or core proteins is formulated with a suitable adjuvant and administered by injection to human volunteers, according to known methods for producing human antisera. Antibody response to the injected proteins is monitored, during a several-week period following immunization, by periodic serum sampling to detect the presence of anti-HCV E1 and/or anti-HCV core serum antibodies, using an immunoassay as described herein.

The antiserum from immunized individuals may be administered as a pre-exposure prophylactic measure for individuals who are at risk of contracting infection. The antiserum is also useful in treating an individual postexposure, analogous to the use of high titer antiserum against hepatitis B virus for post-exposure prophylaxis. For both in vivo use of antibodies to HCV virus-

like particles and proteins and anti-idiotype antibodies and diagnostic use, it may be preferable to use monoclonal antibodies. Monoclonal anti-HCV E1 and anti-HCV core protein antibodies or anti-idiotype antibodies can be produced as follows. The spleen or lymphocytes from an immunized animal are removed and immortalized or used to prepare hybridomas by methods known to those skilled in the (Goding, J.W. 1983. Monoclonal Antibodies: Principles and Practice, Pladermic Press, Inc., NY, NY, pp. 56-97). To produce a human-human hybridoma, a human

lymphocyte donor is selected. A donor known to be infected with HCV (where infection has been shown for example by the presence of anti-virus antibodies in the blood or by virus culture) may serve as a suitable lymphocyte donor.

Lymphocytes can be isolated from a peripheral blood sample

372577 1

5

10

15

20

25

30

or spleen cells may be used if the donor is subject to splenectomy. Epstein-Barr virus (EBV) can be used to immortalize human lymphocytes or a human fusion partner can be used to produce human-human hybridomas. Primary in vitro immunization with peptides can also be used in the generation of human monoclonal antibodies.

Antibodies secreted by the immortalized cells are screened to determine the clones that secrete antibodies of the desired specificity. For monoclonal anti-E1 and anti-core antibodies, the antibodies must bind to HCV E1 and core proteins respectively. For monoclonal anti-idiotype antibodies, the antibodies must bind to anti-E1 and anti-core protein antibodies respectively. Cells producing antibodies of the desired specificity are selected.

The present invention also relates to the use of

single-stranded antisense poly- or oligonucleotides derived 15 from nucleotide sequences substantially homologous to those shown in SEQ ID NOs:1-51 to inhibit the expression of hepatitis C El genes. The present invention further relates to the use of single-stranded anti-sense poly- or oligo-nucleotides derived from nucleotide sequences 20 substantially homologous to those shown in SEO ID NOs:103-154 to inhibit the expression of hepatitis C core genes. Alternatively, the anti-sense poly- or oligo-nucleotides may be complementary to both the E1 and core genes and hence, inhibit the expression of both hepatitis C E1 and 25 core genes. By substantially homologous as used throughout the specification and claims to describe the nucleic acid sequences of the present invention, is meant a level of homology between the nucleic acid sequence and the SEQ ID NOs. referred to in the above sentence. Preferably, the 30 level of homology is in excess of 80%, more preferably in excess of 90%, with a preferred nucleic acid sequence being in excess of 95% homologous with the DNA sequence shown in the indicated SEQ ID NO. These anti-sense poly- or oligonucleotides can be either DNA or RNA. The targeted 35

372577\_1

sequence is typically messenger RNA and more preferably, a single sequence required for processing or translation of the RNA. The anti-sense poly- or oligonucleotides can be conjugated to a polycation such as polylysine as disclosed in Lemaitre, M. et al. ((1989) Proc. Natl. Acad. Sci. USA 84:648-652) and this conjugate can be administrated to a mammal in an amount sufficient to hybridize to and inhibit the function of the messenger RNA.

The present invention further relates to multiple computer-generated alignments of the nucleotide and deduced amino acid sequences shown in SEQ ID NOs:1-206. Computer analysis of the nucleotide sequences shown in SEQ ID NOs:151 and 103-154 and of the deduced amino acid sequences shown in SEQ ID NOs:52-102 and 155-206 can be carried out using commercially available computer programs known to one skilled in the art.

In one embodiment, computer analysis of SEQ ID NOs:1-51 by the program GENALIGN (Intelligenetics, Inc. Mountainview, CA) results in distribution of the 51 HCV E1 sequences into twelve genotypes based upon the degree of variation of the sequences. For the purposes of the present invention, the nucleotide sequence identity of E1 cDNAs of HCV isolates of the same genotype is in the range of about 85% to about 100% whereas the identity of E1 cDNA sequences of different genotypes is in the range of about 50% to about 80%.

The grouping of SEQ ID NOs:1-51 into twelve HCV genotypes is shown below.

30

5

10

15

20

۰	SEQ ID NOs:	Genotypes
5	1-8 9-25 26-29 30-33 34 35-39 40 41 42-43 44 45-50 51	I/1a II/1b III/2a IIV/2b 2c V/3a 4a 4b 4c 4d 5a 6a

For those genotypes containing more than one E1 nucleotide sequence, computer alignment of the constituent nucleotide sequences of the genotype was conducted using GENALIGN in order to produce a consensus sequence for each genotype. These alignments and their resultant consensus sequences are shown in Figures 1A-G for the seven genotypes (I/1a, II/1b, III/2a, IV/2b, V/3a, 4c and 5a) which comprise more than one nucleotide sequence. Further alignment of the consensus sequences of Figures 1A-G with SEQ ID NO:34 (genotype 2c), SEQ ID NO:40 (genotype 4a), SEQ ID NO:41 (genotype 4b), SEQ ID NO:44 (genotype 4d) and SEQ ID NO:51 (genotype 6a) produces a consensus sequence for all twelve genotypes as shown in Figure 1H. The multiple alignments of nucleotide sequences shown in Figures 1A-H produce consensus sequences which serve to highlight regions of homology and non-homology between sequences found within the same genotype or in different genotypes and hence, these alignments can be used by one skilled in the art to design oligonucleotides useful as reagents in

Examples of purified and isolated oligonucleotide sequences derived from the consensus sequences shown in Figures 1A-H include, but are not limited to, SEQ ID NOS:213-239 where these oligonucleotides are useful as "genotype-specific" primers and probes since these

10

15

20

25

30

diagnostic assays for HCV.

oligonucleotides can hybridize specifically to the nucleotide sequence of the E1 gene of HCV isolates belonging to a single genotype. The genotype-specificity of the oligonucleotides shown in SEQ ID NOs:213-239 is as follows: SEQ ID NOs:213-214 are specific for genotype 5 I/la; SEO ID NOs:215-216 are specific for genotype II/lb; SEQ ID NOs:217-218 are specific for genotype III/2a; SEQ ID NOs:219-220 are specific for genotype IV/2b; SEQ ID NOs:221-223 are specific for genotype 2c; SEQ ID NOs:224-226 are specific for genotype V/3a; SEQ ID NOs:227-228 are specific for genotype 4a; SEQ ID NOs:229-230 are specific 10 for genotype 4b; SEQ ID NOs:231-232 are specific for genotype 4c; SEQ ID NOs:233-234 are specific for genotype 4d; SEQ ID NOs:235-236 are specific for genotype 5a and SEQ ID NOs:237-239 are specific for genotype 6a.

In another embodiment, the computer analysis of SEQ ID NOs:103-154 by the program GENALIGN results in distribution of the 52 HCV core sequences into 14 genotypes based upon the degree of variation of the sequences.

The grouping of SEQ ID NOs:103-154 into 14 HCV genotypes is shown below.

	SEQ ID NOs:	<u>Genotypes</u>
25	103-108 109-124 125-128 129-133 134 135-138	I/1a II/1b III/2a IV/2b 2c V/3a 4a
30	141 143 144 145 142 140 146-153	40 40 40 4d 4e 4f 5a 6a

These 14 genotypes can be further grouped into 6

15

major genotypes designated genotypes 1-6 where genotype 1 comprises the sequences contained in minor genotypes I/la and II/1b; genotype 2 comprises the sequences contained in minor genotypes III/2a, IV/2b and 2c; genotype 3 comprises sequences contained in genotype V/3a; genotype 4 comprises 5 sequences contained in minor genotypes 4a-4f; genotype 5 comprises the sequences contained in genotype 5a and genotype 6 comprises the sequence contained in genotype 6a. Computer alignment of the constituent nucleotide sequences of the core cDNAs falling within genotypes I/la, II/lb, III/2a, IV/2b, V/3a and 5a, to produce a consensus sequence 10 for each of these genotypes is shown in Figures 6A (I/la), 6B (II/1b), 6D (III/2a), 6E (IV/2b), 6G (V/3a) and 6I (5a). The alignment of the sequences found in minor genotypes I/la and II/lb to produce a consensus sequence for major genotype 1 is shown in Figure 6C. The alignment of the 15 sequences contained in minor genotypes III/2a, IV/2b and 2c to produce a consensus sequence for major genotype 2 is shown in Figure 6F. The alignment of the nucleotide sequences contained in minor genotypes 4a-4f to produce a consensus sequence for major genotype 4 is shown in Figure 20 Further alignment of the consensus sequences shown in Figures 6C, 6F, 6G, 6H and 6I with SEQ ID NO:154 (genotype 6a/major genotype 6) to produce a consensus sequence for all genotypes is shown in Figure 6J and alignment of the consensus sequences shown in Figures 6A, 6B, 6D, 6E, 6G and 25 6I with 4a), SEQ ID NO:141 (genotype 4b), SEQ ID NO:143 (genotype 4c), SEO ID NO:145 (genotype 4d), SEO ID NO:142 (genotype 4e), SEQ ID NO:140 (genotype 4f) and SEQ ID NO:154 (genotype 6a) to produce a consensus sequence for all fourteen genotypes is shown in Figure 6K. As with the 30 alignments of the envelope (E1) nucleotide sequences, the consensus sequences shown in Figures 6A-6K serve to highlight regions of homology and non-homology between sequences found within the same genotype or in different genotypes and hence, can be used by one skilled in the art 35 372577 1

10

to design oligonucleotides useful as reagents in diagnostic assays for HCV.

For example, purified and isolated oligonucleotide sequences derived from the consensus sequences shown in Figures 6A-6K may be useful as genotype-specific primers and probes since these oligonucleotides can hybridize specifically to the nucleotide sequence of the core gene of HCV isolates belonging to a given genotype. Examples of regions of the consensus sequence of the core gene of a given genotype from which primers specific for that genotype may be deduced include but are not limited to, the nucleotide domains shown below for each genotype. The sequence in which the indicated nucleotide domains are found are indicated in parentheses to the right

15 Genotype 1 (Consensus Sequence of Figure 6C)
 427-466, 444-483, 447-486 (5'-3', sense)
 505-466, 522-483, 525-486 (5'-3', antisense)

# Genotype 1a (Consensus Sequence of Figure 6A)

20 141-180, 279-318 (5'-3', sense) 219-180, 246-207 (5'-3', antisense)

231, 193-232, 302-341 (5'-3', sense)

of each genotype.

372577 1

# <u>Genotype 1b (Consensus Sequence of Figure 6B)</u> 67-106, 127-186, 234-273 (5'-3', sense)

25 144-106, 225-186, 311-272, 312-273 (5'-3', antisense)

<u>Genotype 2 (Consensus Sequence of Figure 6F)</u> 153-192, 162-201, 164-203, 168-207, 171-210, 182-221, 192-

30 231-192, 240-201, 242-203, 246-207, 249-210, 260-221, 270-231, 271-232, 380-341 (5'-3', antisense)

Genotype III/2a (Consensus Sequence of Figure 6D) 276-315, 306-355 (5'-3', sense)

35 309-270, 354-315, 394-355, 571-532 (5'-3', antisense)

Genotype IV/2b (Consensus Sequence of Figure 6E)

6-45, 135-174, 177-216, 309-348, 337-376, 375-414, 501-540 (5'-3', sense)

84-45, 213-174, 255-216, 387-348, 415-376, 453-414, 571-532, 573-540 (5'-3', antisense)

5

15

#### Genotype 2c (SEQ ID NO:134)

194-233, 273-312, 279-318, 417-456, 423-462, 504-543, 505-544, 517-556 (5'-3', sense)

272-233, 351-312, 354-315, 357-318, 450-411, 495-456, 501-

10 462, 573-543, 556-573 (5'-3', antisense)

# Genotype 3 or Genotype V/3a (Consensus Sequence of Figure 6G)

8-47, 45-84, 68-107, 87-126, 88-127, 90-129, 111-150, 142-181, 173-212, 177-216, 261-300,

276-315, 452-491, 520-559, 521-560, 529-568, 532-571, 533-572. (5'-3', sense)

86-47, 123-84, 146-107, 165-126, 186-147, 189-150, 219-180, 250-211, 251-212, 255-216,

20 339-300, 530-491, 573-543, 573-557, 573-559, 573-560. (5'-3', antisense)

# Genotype 4 (Consensus Sequence of Figure 6H)

20-59 (5'-3', sense)

25 97-58, 98-59 (5'-3', antisense)

# Genotype 4a (SEQ ID NO:139)

111-150, 150-189, 174-213, 183-222, 192-231, 261-300, 376-415, 396-435, 531-570 (5'-3', sense)

30 186-147, 252-213, 270 -231, 339-300, 454-415 (5'-3', antisense)

# Genotype 4b (SEQ ID NO:141)

27-66, 30-69, 106-145, 271-310, 433-472, 447-486, 453-492 (5'-3', sense)

372577\_1

° 105-66, 183-144, 184-145, 345-306, 348-309, 349-310, 468-429, 510-471, 522-483, 570-531 (5'-3', antisense)

## Genotype 4c (SEQ ID NO:143

174-213, 180-219, 207-246, 231-270 (5'-3', sense)

5 249-210, 252-213, 258-219, 309-270, 504-465 (5'-3', antisense)

# Genotype 4d (SEQ ID NO:145)

173-212, 188-327, 430-469 (5'-3', sense)

10 248-209, 249-210, 250-211, 251-212, 366-327, 508-469 (5'-3', antisense)

## Genotype 4e (SEQ ID NO:142)

160-199, 267-306, 287-326, 288-327, 524-564 (5'-3', sense)

15 238-199, 345-306, 365-326, 216-177, 522-483 (5'-3', antisense)

### Genotype 4f (SEQ ID NO:140)

18-57, 36-75, 228-267, 396-435 (5'-3', sense)

20 96-57, 114-75, 306-267 (5'-3', antisense)

#### Genotype 5 or 5a (Consensus Sequence of Figure 61)

176-215, 177-216, 181-220, 195-234, 221-260, 252-291, 255-294, 396-435, 435-474, 447-486, 498-537 (5'-3', sense)

25 254-215, 299-260, 310-271, 330-291, 333-294, 354-315, 464-425, 471-432, 483-444, 570-531 (5'-3', antisense)

### Genotype 6 or 6a (SEQ ID NO:154)

20-59, 136-175, 156-195, 159-198, 175-214, 185-224, 277-

316, 278-317, 312-351, 348-387,405-444, 406-445, 407-446, 408-447, 411-450, 432-471, 433-472, 435-474, 522-561 (5'-3', sense).

98-59, 214-175, 234-195, 237-198, 253-214, 262-223, 263-224, 354-315, 355-316, 382-343, 390-351, 426-387, 468-429,

35 483-444, 484-445, 485-446, 486-447, 489-450, 510-471, 511-

3. sense); and

472, 513-474 (5'-3', antisense)

Such nucleotide domains may range from about 15 to about 100 bases in length with a more preferred range being about 30 to about 60 bases in length.

In an alternative embodiment, universal primers

able to hybridize to the nucleotide sequences of the core
gene of HCV isolates belonging to all of the genotypes
disclosed herein may be deduced from universally conserved
nucleotide domains of the consensus sequence shown in
Figures 6J and 6K. Examples of such nucleotide domains
include, but are not limited to, those shown below:

nucleotides 1-20, 1-25, 1-26, 1-27, 1-33, 50-89, 51-90, 52-91, 53-92, 61-100, 62-101, 77-116, 78-117, 79-118, 80-119, 81-120, 82-121, 83-122, 84-123, 85-124, 86-125, 97-136, 98-137, 99-138, 100-139, 101-140, 102-141, 329-368, 330-369, 331-370, 332-371, 354-393, 355-394, 356-395, 362-401, 363-402, 364-403, 365-404, 369-408, 442-481, 443-482, 457-496, 458-497, 475-514, 476-515, 477-516 (5'-

nucleotides 40-1, 41-2, 42-3, 43-4, 51-12, 52-13, 55-16, 56-17, 57-18, 58-19, 61-22, 62-23, 63-24,

20 55-16, 56-17, 57-18, 58-19, 61-22, 62-23, 63-24, 64-25, 70-31, 124-85, 125-86, 126-87, 127-88, 128-89, 129-90, 136-97, 137-98, 138-99, 149-110, 150-111, 151-112, 152-113, 153-114, 154-115, 155-

116, 156-117, 157-118, 158-119, 159-120, 170-131, 171-132, 172-133, 173-134, 174-135, 175-136, 403-364, 405-365, 406-

25 172-133, 173-134, 174-135, 175-136, 403-364, 405-365, 406-366, 406-367, 430-391, 431-392, 432-393, 436-397, 437-398, 438-399, 439-400, 517-478, 518-479, 519-480, 532-493, 533-494, 550-511, 551-512 (5'-3', antisense)

Those skilled in the art would readily understand that the term "antisense" as used herein refers to primer

sequences which are the complementary sequence of the indicated consensus sequence or SEQ ID NO:. Further, provided with the above examples of regions of the consensus sequences or indicated SEQ ID NOS: from which to

deduce universal and genotype-specific primers, those

372577 1

10

skilled in the art would readily be able to select pairs of primers, one sense and one antisense, which would be useful in the detection of HCV genotypes via the PCR methods described herein.

In yet another embodiment, the sequences shown in SEQ ID NO.:103-154 and the resultant consensus sequences produced by alignment of these SEQ ID NOs as shown in Figures 6A-6K may also be useful in the design of hybridization probes specific for a given HCV genotype. Examples of nucleotide domains of the consensus sequence or SEQ ID NO of a given genotype from which genotype-specific hybridization probes may be deduced include, but are not limited to, those shown below where the sequence from which the domains are found is indicated in parentheses to the right of each genotype.

15	_	Ger	notype	Positio				
	1a	(Consensus	sequence	of	Figure	6A)	50-85 155-205 207-277 281-333 429-477 530-573	
20	1b	(Consensus	sequence	of	Figure	6B)	81-131 159-225 252-318 411-472 530-573	
25	2a	(Consensus	sequence	of	Figure	6D)	35-75 200-276 290-340 330-380 410-472 530-573	
30	2b	(Consensus	sequence	of	Figure	6E)	20-70 149-199 191-241 240-285 261-318 323-373 351-401 389-439	
35							429-477	

	- 44 -	
٥		530-573
5	2c (SEQ ID NO:134)	208-258 230-276 290-345 411-460 430-490 530-573
10	3a (Consensus sequence of Figure 6G)	1-50 40-100 100-160 145-190 190-240 275-325 411-455 466-516 530-573
15	4a (SEQ ID NO:139)	35-85 145-195 200-250 255-305 341-390 390-440 530-573
20	4b (SEQ ID NO:141)	35-85 120-170 180-225 230-275 285-335 405-455 462-492 530-573
25	4c (SEQ ID NO:143)	35-85 190-246 245-295 282-318 372-415 440-480 530-573
30	4d (SEQ ID NO:145)	35-85 187-237 302-352 405-455 444-494 530-573
35	4e (SEQ ID NO:142)	35-85 57-84

۰			174-224 230-275 290-340 422-472 530-573
5		4f (SEQ ID NO:140)	35-85 174-224 242-292 290-340 422-472 530-573
10	5a	(Consensus sequence of Figure 61)	180-234 265-315 315-355 420-486 530-573
15		6a (SEQ ID NO:154)	34-84 150-200 180-230 230-290 291-333 341-395 429-490 530-573
	1	(Consensus sequence of Figure 6C)	192-241 435-495
20	2	(Consensus sequence of Figure 6F)	186-240 320-360 440-475
	4	(Consensus sequence of Figure 6H)	40-80

In yet another embodiment, universal hybridization probes may be derived from the consensus sequences shown in Figures 6J and 6K. Examples of nucleotide domains of the consensus sequences shown in Figure 6J and 6K from which universal hybridization probes may be derived include, but are not limited to, 1-33; 85-141; 364-408; 478-516.

The oligonucleotides of this invention can be synthesized using any of the known methods of oligonucleotide synthesis (e.g., the phosphodiester method of Agarwal et al. 1972, Agnew. Chem. Int. Ed. Engl. 11:451,

372577\_1

25

30

the phosphotriester method of Hsiung et al. 1979, Nucleic Acids Res 6:1371, or the automated diethylphosphoramidite method of Baeucage et al. 1981, Tetrahedron Letters 22:1859-1862), or they can be isolated fragments of naturally occurring or cloned DNA. In addition, those skilled in the art would be aware that oligonucleotides can 5 be synthesized by automated instruments sold by a variety of manufacturers or can be commercially custom ordered and prepared. In a preferred embodiment, the oligonucleotides of the present invention are synthetic oligonucleotides. The oligonucleotides of the present invention may range 10 from about 15 to about 100 nucleotides; with the preferred sizes being about 20 to about 60 nucleotides; a more preferred size being about 25 to about 50 nucleotides; and

a most preferred size being about 30 to about 40

The present invention also relates to methods for detecting the presence of HCV in a mammal, said methods comprising analyzing the RNA of a mammal for the presence of hepatitis C virus.

The RNA to be analyzed can be isolated from 20 serum, liver, saliva, lymphocytes or other mononuclear cells as viral RNA, whole cell RNA or as poly(A) $^+$  RNA. Whole cell RNA can be isolated by methods known to those skilled in the art. Such methods include extraction of RNA by differential precipitation (Birnbiom, H.C. (1988) 25 Nucleic Acids Res., 16:1487-1497), extraction of RNA by organic solvents (Chomczynski, P. et al. (1987) Anal. Biochem., 162:156-159) and extraction of RNA with strong denaturants (Chirgwin, J.M. et al. (1979) Biochemistry, 18:5294-5299). Poly(A) + RNA can be selected from whole cell 30 RNA by affinity chromatography on oligo-d(T) columns (Aviv, H. et al. (1972) Proc. Natl. Acad. Sci., 69:1408-1412). A preferred method of isolating RNA is extraction of viral RNA by the guanidinium-phenol-chloroform method of Bukh et al. (1992a). 35

372577 1

nucleotides.

15

20

25

30

35

The methods for analyzing the RNA for the presence of HCV include Northern blotting (Alwine, J.C. et al. (1977) Proc. Natl. Acad. Sci., 74:5350-5354), dot and slot hybridization (Kafatos, F.C. et al. (1979) Nucleic Acids Res., 7:1541-1522), filter hybridization (Hollander, M.C. et al. (1990) Biotechniques; 9:174-179), RNase protection (Sambrook, J. et al. (1989) in "Molecular Cloning, A Laboratory Manual", Cold Spring Harbor Press, Plainview, NY) and reverse-transcription polymerase chain reaction (RT-PCR) (Watson, J.D. et al. (1992) in

10 "Recombinant DNA" Second Edition, W.H. Freeman and Company, New York).

A preferred method for analyzing the RNA is RT-

In this method, the RNA can be reverse transcribed to first strand cDNA using a primer or primers derived from the nucleotide sequences shown in SEQ ID NOs:1-51 or SEQ ID NOs:103-154 or sequences complementary to those described. Once the cDNAs are synthesized, PCR amplification is carried out using pairs of primers designed to hybridize with sequences in the HCV E1 or core cDNA which are an appropriate distance apart (at least about 50 nucleotides) to permit amplification of the cDNA and subsequent detection of the amplification product. Alternatively, one can amplify both E1 and core cDNA sequences by using a primer pair where one primer hybridizes with the E1 cDNA sequence and the other primer hybridizes with the core cDNA sequence. Each primer of a pair is a single-stranded oligonucleotide of about 20 to about 60 bases in length with a more preferred range being about 30 to about 50 bases in length where one primer (the "upstream" primer) is complementary to the original RNA and the second primer (the "downstream" primer) is complementary to the first strand of cDNA generated by reverse transcription of the The target sequence is generally about 100 to about 300 base pairs long but can be as large as 500-1500 base pairs. Optimization of the amplification reaction to

372577\_1

10

15

20

2.5

30

35

obtain sufficiently specific hybridization to the nucleotide sequence of interest (either E1 or core or both E1 and core) is well within the skill in the art and is preferably achieved by adjusting the annealing temperature.

In one embodiment, the primer pairs selected to amplify E1 and core cDNAs are universal primers. "universal", as used to describe primers throughout the claims and specification, is meant those primer pairs which can amplify E1 and/or core gene fragments derived from an HCV isolate belonging to any one of the genotypes of HCV described herein. Purified and isolated universal primers for E1 cDNAs are used in Example 1 of the present invention and are shown as SEQ ID NOs:207-212 where SEQ ID NOs:207 and 208 represent one pair of primers, SEQ ID NOs:209 and 210 represent a second pair of primers and SEQ ID NOs:211-212 represent a third pair of primers. Nucleotide domains of the consensus sequence shown in Figure 6J from which universal primers for core cDNAs may be deduced have previously been disclosed within the present specification. Alternatively, a universal primer for El cDNA sequence and a universal primer for core cDNA sequence may be used as a universal primer pair to amplify both E1 and core cDNAs.

In an alternative embodiment, primer pairs selected to amplify E1 and/or core cDNAs are genotype-specific primers. In the present invention, genotype-specific primer pairs can readily be derived from the following genotype-specific E1 nucleotide domains: nucleotides 197-238 and 450-480 of the consensus sequence of genotype I/la shown in Figure 1A; nucleotides 197-238 and 450-480 of the consensus sequence of genotype II/lb shown in Figure 1B; nucleotides 199-238 and 438-480 of the consensus sequence of genotype III/2a shown in Figure C; nucleotides 124-177 and 450-480 of the consensus sequence of genotype IV/2b shown in Figure 1D; nucleotides 124-177, 193-238 and 436-480 of SEQ ID NO:34 (genotype 2C); nucleotides 168-207, 294-339 and 406-480 of the consensus

sequence of genotype V/3a shown in Figure 1E; nucleotides 145-183 and 439-480 of SEQ ID NO:40 (genotype 4a); nucleotides 168-207 and 432-480 of SEQ ID NO:41 (genotype 4b); nucleotides 130-183 and 450-480 of the consensus sequence of genotype 4c shown in Figure 1F; nucleotides 130-183 and 450-480 of SEQ ID NO:44 (genotype 4d); 5 nucleotides 166-208 and 437-480 of the consensus sequence of genotype 5a shown in Figure 1b and nucleotides 168-207, 216-252 and 429-480 of SEO ID NO:51 (genotype 6a). Genotype-specific HCV core nucleotide domains from which genotype-specific primers may be deduced have previously 10 been described herein. Those skilled in the art would readily appreciate that in a pair of genotype-specific primers, each primer is derived from different nucleotide domains specific for a given genotype. Also, it is understood by those skilled in the art that each pair of 15 primers comprises one primer which is complementary to the original viral RNA and the other which is complementary to the first strand of cDNA generated by reverse transcription of the viral RNA. For example, in a pair of genotypespecific primers for genotype 4b, one primer would have a 20 nucleotide sequence derived from region 168-207 of SEQ ID

NO:40 and the other primer would have a nucleotide sequence which is the complement of region 432-480 of SEQ ID NO:40. One skilled in the art would readily recognize that such genotype-specific domains would also be useful in designing oligonucleotides for use as genotype-specific hybridization probes. Indeed, genotype-specific hybridization probes deduced from the E1 and core sequences of the present invention have been previously disclosed herein.

The amplification products of PCR can be detected either directly or indirectly. In one embodiment, direct detection of the amplification products is carried out via labelling of primer pairs. Labels suitable for labelling the primers of the present invention are known to one skilled in the art and include radioactive labels, biotin,

372577 1

25

30

25

30

35

avidin, enzymes and fluorescent molecules. The derived labels can be incorporated into the primers prior to performing the amplification reaction. A preferred labelling procedure utilizes radiolabeled ATP and T4 polynucleotide kinase (Sambrook, J. et al. (1989) in "Molecular Cloning, A Laboratory Manual", Cold Spring 5 Harbor Press, Plainview, NY). Alternatively, the desired label can be incorporated into the primer extension products during the amplification reaction in the form of one or more labelled dNTPs. In the present invention, the labelled amplified PCR products can be detected by agarose 10 gel electrophoresis followed by ethidum bromide staining and visualization under ultraviolet light or via direct sequencing of the PCR-products. Thus, in one embodiment, the present invention relates to a method for determining the genotype of a hepatitis C virus present in a mammal 15 where said method comprises: amplifying RNA of a mammal via RT-PCR using labelled genotype-specific primers for the amplification step of the cDNA produced by reverse

In yet another embodiment, unlabelled amplification products can be detected via hybridization with labelled nucleic acid probes radioactively labelled or, labelled with biotin, in methods known to one skilled in the art such as dot and slot blot hybridization (Kafatos, F.C. et al. (1979) or filter hybridization (Hollander, M.C. et al. (1990)).

In one embodiment, the nucleic acid sequences used as probes are selected from, and substantially homologous to, SEQ ID NOs:1-51 and/or SEQ ID NOs:103-154. Such probes are useful as universal probes in that they can detect PCR-amplification products of E1 and/or core cDNAs of an HCV isolate belonging to any of the HCV genotypes disclosed herein. The size of these probes can range from about 200 to about 500 nucleotides. In an alternative embodiment, the sequence alignments shown in Figures 1A-1H

372577\_1

transcription.

10

15

20

25

30

35

and 6A-6J may be used to design oligonucleotides useful as universal hybridization probes. Examples of core and envelope nucleotide domains from which such universal oligonucleotides may be deduced are disclosed herein.

In yet another embodiment, the present invention relates to a method for determining the genotype of a hepatitis C virus present in a mammal where said method comprises:

- (a) amplifying RNA of a mammal via RT-PCR to produce amplification products;
  - (b) contacting said products with at least one genotype-specific oligonucleotide; and
    - (c) detecting complexes of said products which bind to said oligonucleotide(s).

In this method, one embodiment of said amplification step is carried out using the universal primers for E1 or core cDNAs as disclosed above. In step (b) of this method, the genotype-specific sequences used as probes may be deduced from the genotype-specific E1 and core nucleotide domains disclosed herein. These probes are useful in specifically detecting PCR-amplification products of E1 or core cDNAs of HCV isolates belonging to one of the HCV genotypes disclosed herein. In a preferred embodiment, these probes are used alone or in combination with other probes specific to the same genotype.

For example, a probe having a sequence according to SEQ ID NO:213 can be used alone or in combination with a probe having a sequence according to SEQ ID NO:214. The probes used in this method can range in size from about 15 to about 100 nucleotides with a more preferred range being about 30 to about 70 nucleotides. Such probes can be synthesized as described earlier.

In an alternative embodiment, the genotype of the amplification product of step (a) may be determined by using the nucleic acid sequences shown in SEQ ID NOs: 1-51 and 103-154 as probes (Delwart, E. et al. (1993)) <a href="Science">Science</a>,

10

15

20

25

30

35

262: 1257-1261). Probes utilized in the method of Delwart et al. may range in size from about 100 to about 1,000 nucleotides with a more preferred probe size being about 200 to about 800 base pairs and a most preferred probe size being about 300 to about 700 nucleotides.

The nucleic acid sequence used as a probe to detect PCR amplification products of the present invention can be labeled in single-stranded or double-stranded form. Labelling of the nucleic acid sequence can be carried out by techniques known to one skilled in the art. labelling techniques can include radiolabels and enzymes (Sambrook, J. et al. (1989) in "Molecular Cloning, A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In addition, there are known non-radioactive techniques for signal amplification including methods for attaching chemical moieties to pyrimidine and purine rings (Dale, R.N.K. et al. (1973) Proc. Natl. Acad. Sci., 70:2238-2242; Heck, R.F. (1968) S. Am. Chem. Soc., 90:5518-5523), methods which allow detection by chemiluminescence (Barton, S.K. et al. (1992) J. Am. Chem. Soc., 114:8736-8740) and methods utilizing biotinylated nucleic acid probes (Johnson, T.K. et al. (1983) Anal. Biochem., 133:126-131; Erickson, P.F. et al. (1982) J. of Immunology Methods, 51:241-249; Matthaei, F.S. et al. (1986) Anal.

Blochem., 157:123-128) and methods which allow detection by fluorescence using commercially available products.

The present invention also relates to computer analysis of the amino acid sequences shown in SEQ ID NOs:52-102 by the program GENALIGN. This analysis groups the 51 amino acid sequences shown in SEQ ID NOs:52-102 into twelve genotypes based upon the degree of variation of the amino acid sequences. For the purposes of the present invention, the amino acid sequence identity of El amino acid sequences of the same genotype ranges from about 85% to about 100% whereas the identity of E1 amino acid sequences of different genotypes ranges from about 45% to

about 80%.

The grouping of SEQ ID NOs:52-102 into twelve HCV genotypes is shown below:

	SEQ ID NOs:	<u>Genotypes</u>
5	52-59 60-76 77-80 81-84 85	I/1a II/1b III/2a IV/2b 2c
10	86-90 91 92 93-94 95 96-101	20 4a 4b 4c 4d 5a 6a

For those genotypes containing more than one E1 amino acid sequence, computer alignment of the constituent sequences of each genotype was conducted using the computer program GENALIGN in order to produce a consensus sequence for each genotype. These alignments and their resultant consensus sequences are shown in Figures 2A-G for the seven genotypes (I/la, II/lb, III/2a, IV/2b, V/3a, 4c and 5a) which comprise more than one sequence. Further alignment of the consensus sequences shown in Figures 2A-G with the amino acid sequences of SEQ ID NO:85 (genotype 2c); SEQ ID NO:91 (genotype 4a); SEQ ID NO:92 (genotype 4b); SEQ ID NO:95 (genotype 4d) and SEQ ID NO:102 (genotype 6a) to produce a consensus amino acid sequence for all twelve genotypes is shown in Figure 2H. The multiple alignment of E1 amino acid sequences shown in Figures 2A-H produces consensus sequences which serve to highlight regions of homology and non-homology between E1 amino acid sequences of the same genotype and of different genotypes and hence, these alignments can readily be used by those skilled in the art to design peptides useful in assays and vaccines for the diagnosis and prevention of HCV infection.

In another embodiment, the computer analysis of

35

15

20

25

SEQ ID NOS: 155-206 by the probe genome results in distribution of the 52 HCV core sequences into 14 genotypes based upon identification of genotype-specific amino acid sequences.

The grouping of SEQ ID NOS: 155-206 into 14 HCV genotypes is shown below:

	SEO ID NOS:	Genotypes
	155-160 161-176	I/1a II/1b
10	177-180	III/2a
	181-185	IV/2b
10	186	2c
	187-190	V/3a
	191	4a
	193	4b
	195	4 C
	196	4 C
1.5	197	4d
15	194	4 e
	192	4 f
	198-205	5a
	206	6a

These fourteen genotypes can be further grouped into six major genotypes designated genotypes 1-6 as described earlier for the core nucleotide sequences of the present application. Computer alignment of the amino acid sequences disclosed in SEQ ID NOS: 155-206 are shown in figures 7A-7J. As with the multiple alignments of the E-1 amino acid sequences, the consensus sequences shown in figure 7A-7J serve to highlight regions of homology and nonhomology between core amino acid sequences of the same genotype and of different genotypes and hence, these alignments can readily be used by those skilled in the art to design peptides useful in assays and vaccines for the diagnosis and prevention of HCV infection.

Examples of purified and isolated peptides deduced from the alignments shown in Figures 2A-2H include, but are not limited to, SEQ ID NOs:240-263 wherein these peptides are derived from two regions of the amino acid

35

5

20

25

20

25

30

35

sequences shown in Figures 2A-H, amino acids 48-80 and amino acids 138-160. The peptides shown in SEQ ID NOs. 240-263 are useful as genotype-specific diagnostic reagents since they are capable of detecting an immune response specific to HCV isolates belonging to a single genotype. 5 The genotype-specificity of the peptides shown in SEQ ID NOs:240-263 are as follows: SEQ ID NOs:240 and 252 are specific for genotype IV/2b; SEQ ID NOs:241 and 253 are specific for genotype 2c; SEQ ID NOs:242 and 254 are specific for genotype III/2a; SEQ ID NOs:243 and 255 are specific for genotype V/a; SEQ ID NOs:244 and 256 are 10 specific for genotype II/1b; SEQ ID NOs:245 and 257 are specific for genotype I/la; SEQ ID NOs:246 and 258 are specific for genotype 4a; SEQ ID NOs:247 and 259 are

specific for genotype 4c; SEQ ID NOs:248 and 260 are

specific for genotype 4d; SEQ ID NOs:249 and 261 are specific for genotype 4b; SEQ ID NOs:250 and 262 are specific for genotype 5a and SEQ ID NOs:251 and 263 are specific for genotype 6a. In SEQ ID NO:240, Xaa at position 22 is a residue of Ala or Thr, Xaa at position 24 is a residue of Val or Ile, Xaa at position 26 is a residue of Val or Met; in SEQ ID NO:242, Xaa at position 5 is a Ser or Thr residue, Xaa at position 11 is an Arg or Gln residue, Xaa at position 12 is an Arg or Gln residue; in SEQ ID NO:243, Xaa at position 3 is a Pro or Ser residue, Xaa at position 33 is a Leu or Met residue; in SEQ ID NO:244, Xaa at position 5 is a Thr or Ala residue, Xaa at

position 13 is a Gly, Ala, Ser, Val or Thr residue, Xaa at position 14 is a Ser, Thr or Asn residue, Xaa at position 15 is a Val or Ile residue, Xaa at position 16 is a Pro or Ser residue, Xaa at position 18 is a Thr or Lys residue, Xaa at position 19 is a Thr or Ala residue, Xaa at position 22 is an Arg or His residue, Xaa at position 32 is an Ala, Val or Thr residue; in SEQ ID NO:245, Xaa at position 3 is an Ala or Pro residue, Xaa at position 4 is a Val or Met residue, Xaa at position 5 is a Thr or Ala residue, Xaa at

25

30

35

position 17 is a Thr or Ala residue, Xaa at position 18 is a Thr or Ala residue, Xaa at position 23 is a His or Tyr residue; in SEQ ID NO:247, Xaa at position 10 is a Val or Ala residue, Xaa at position 11 is a Ser or Pro residue, Xaa at position 18 is an Asp or Glu residue Xaa at position 20 is a Leu or Ile residue; in SEQ ID NO:250, Xaa at 5 position 3 is a Gln or His residue, Xaa at position 12 is an Asn, Ser or Thr residue, Xaa at position 13 is a Leu or Phe residue, Xaa at position 23 is an Ala or Val residue; in SEQ ID NO:252, Xaa at position 16 is a Val or Ala residue, Xaa at position 18 is a Glu or Gln residue; in SEQ 10 ID NO:254, Xaa at position 2 is an Ala or Thr residue, Xaa at position 4 is a Met or Leu residue, Xaa at position 9 is an Ala or Val residue, Xaa at position 17 is an Ile or Leu residue, Xaa at position 20 is an Ile or Val residue, Xaa at position 21 is a Ser or Gly residue; in SEQ ID NO:151, 15 Xaa at position 9 is a Val or Ile residue, Xaa at position 16 is a Leu or Val residue, Xaa at position 20 is an Ile or

Leu residue; in SEQ ID NO:256, Xaa at position 2 is an Ala or Thr residue, Xaa at position 6 is a Val or Leu residue, Xaa at position 12 is an Ile or Leu residue, Xaa at position 16 is a Val or Ile residue, Xaa at position 17 is a Val, Leu or Met residue, Xaa at position 19 is a Met or Val residue, Xaa at position 21 is an Ala or Thr residue; in SEQ ID NO:257, Xaa at position 2 is a Thr or Ala residue, Xaa at position 6 is a Val, Ile or Met residue, Xaa at position 12 is an Ile or Val residue, Xaa at position 16 is a Ile or Val residue; in SEQ ID NO:155, Xaa

at position 5 is a Leu or Val residue, Xaa at position 21 is a Thr or Ala residue; in SEQ ID NO:262, Xaa at position 1 is a Thr or Ala residue, Xaa at position 5 is a Val or Leu residue, Xaa at position 9 is a Leu, Met or Val residue, Xaa at position 23 is a Gly or Ala residue.

Examples of core amino acid domains from which genotype-specific peptides may be deduced, include but are not limited to, those shown below where the sequence in

which the indicated domains are found is given in parentheses to the right of each genotype:

	<u>Genotype</u>	Amino Acid Domains
1a 1b 2	(consensus sequence of Figure 7A) (consensus sequence of Figure 7B) (consensus sequence of Figure 7F)	67-78 67-78 66-81 110-119
2a	(consensus sequence of Figure 7D)	67-78 115-125
2b	(consensus sequence of Figure 7E)	67-78 123-133
2c	(SEQ ID NO:186)	67-78 75-81
3a	(consensus sequence of Figure 7G)	184-191 8-22 32-46 67-78
4 4a 4b	(consensus sequence of Figure 7H) (SEQ ID NO:191) (SEQ ID NO:193)	158-170 180-191 14-23 67-78 45-57 67-78
4c 4d 4e 4f 5a 6a	(SEQ ID NO:195) (SEQ ID NO:197) (SEQ ID NO:194) (SEQ ID NO:192) (consensus sequence of Figure 7J) (SEQ ID NO:206)	67-78 67-78 67-78 67-78 67-78 67-78 101-108 144-155 157-163
	1b 2 2a 2b 2c 3a 4 4a 4b 4c 4dd 4e 4ff 5a	1a (consensus sequence of Figure 7A) 1b (consensus sequence of Figure 7B) 2 (consensus sequence of Figure 7T) 2a (consensus sequence of Figure 7D) 2b (consensus sequence of Figure 7E) 2c (SEQ ID NO:186)  3a (consensus sequence of Figure 7G)  4 (consensus sequence of Figure 7G)  4 (SEQ ID NO:191) 4b (SEQ ID NO:193)  4c (SEQ ID NO:195) 4d (SEQ ID NO:197) 4e (SEQ ID NO:197) 4e (SEQ ID NO:192) 5a (consensus sequence of Figure 7J)

Those skilled in the art would be aware that the peptides of the present invention or analogs thereof can be synthesized by automated instruments sold by a variety of manufacturers or can be commercially custom-ordered and prepared. The term analog has been described earlier in the specification and for purposes of describing the peptides of the present invention, analogs can further include branched, cyclic or other non-linear arrangements of the peptide sequences of the present invention.

Alternatively, peptides can be expressed from nucleic acid sequences where such sequences can be DNA, cDNA, RNA or any variant thereof which is capable of

25

directing protein synthesis. In one embodiment, restriction digest fragments containing a coding sequence for a peptide can be inserted into a suitable expression vector that functions in prokaryotic or eukaryotic cells. Such restriction digest fragments may be obtained from clones isolated from prokaryotic or eukaryotic sources which encode the peptide sequence.

Suitable expression vectors and methods of isolating clones encoding the peptide sequences of the present invention have previously been described. In yet another embodiment, an oligonucleotide capable of directing host organism synthesis of the given peptide may be synthesized and inserted into the expression vector.

The preferred size of the peptides of the present invention is from about 8 to about 100 amino acids in length when the peptides are chemically synthesized with a more preferred size being about 8 to about 30 amino acids and a most preferred size being about 10 to about 20 amino acids in length. For recombinantly expressed peptides, the size may range from about 20 to about 190 amino acids in length with a more preferred size being about 70 amino acids.

The present invention further relates to the use of genotype-specific peptides in methods of detecting antibodies against a specific genotype of HCV in biological samples. In one embodiment, at least one genotype-specific peptide deduced from a genotype-specific core or E1 amino acid domain may be used in any of immunoassays described herein to detect antibodies specific for a single genotype of HCV. In another embodiment, at least one genotype-specific peptide deduced from a genotype-specific core nucleotide domain and at least one genotype-specific peptide deduced from an E1 amino acid domain may be used in an immunoassay to detect antibodies against a single genotype of HCV. A preferred immunoassay is ELISA.

It is understood by those skilled in the art that

35

10

15

20

25

10

15

20

the diagnostic assays described herein using genotypespecific oligonucleotides or genotype-specific peptides can be useful in assisting one skilled in the art to choose a course of therapy for the HCV-infected individual.

In an alternative embodiment, a mixture of genotype-specific peptides can be used in an immunoassay to detect antibodies against multiple genotypes of HCV disclosed herein. For example, a mixture of genotypespecific peptides deduced from E1 amino acid sequences may comprise at least one peptide selected from SEQ ID NOs:244-245 and 256-257; one peptide selected from SEQ ID NOs:240, 242, 252 and 254; one peptide selected from SEQ ID NOs:246-249 and 258-261; one peptide selected from SEQ ID NOs:250 and 262; one peptide selected from SEQ ID NOs:243 and 255; one peptide selected from SEO ID NOs:242 and 254 and one peptide selected from SEQ ID NOs:244 and 263. In a preferred embodiment, the peptides of the present invention can be used in an ELISA assay as described previously for recombinant E1 and core proteins.

In an alternative embodiment, the peptide(s) utilized in an immunoassay to detect all the genotypes of HCV disclosed herein may be a universal peptide deduced from universally conserved amino acid domains of the E1 or core proteins disclosed herein.

Examples of universally conserved core amino acid domains within the consensus sequence shown in Figure 7J 25 from which universal peptides may be deduced include, but are not limited to amino acid domains 23-35, 53-66, 93-108, 122-138, 150-156, and 165-181 of the consensus sequence. Examples of universally conserved E1 amino acid domains

within the HCV E1 protein are located within the consensus 30 sequence for the 51 HCV E1 proteins shown in Figure 2H of the present application. Examples of universally conserved domains within the consensus sequence shown in Figure 2H include, but are not limited to, amino acid domains 10-20, 111-120, and 124-137 of the consensus sequence. 35

15

20

25

universal peptides of the present invention may be used in an immunoassay to detect antibodies in patient sera specific for any of the genotypes of HCV disclosed herein.

The peptides of the present invention or analogs thereof may be prepared in the form of a kit, alone or in combinations with other reagents such as secondary antibodies, for use in immunoassay.

In another embodiment, the genotype-specific and universal peptides of the present invention may be used to produce antibodies that will react against HCV E1 or core proteins in immunoassays. In one embodiment, a genotype-specific E1 or core peptide can be used alone or in combination with other E1 or core peptides specific to the same genotype as immunogens to produce antibodies specific to HCV proteins of a single genotype.

In another embodiment, a mixture of peptides specific for different genotypes may be used to produce antibodies that will react with HCV proteins of any genotype disclosed herein. More preferably, antibodies reactive with HCV proteins of any genotype may be produced by immunizing an animal with universal peptide(s) of the present invention. Examples of immunoassays in which such antibodies could be utilized to detect HCV E1 and core proteins in biological samples include, but are not limited to, radioimmunoassays and ELISAs. Examples of biological samples in which HCV E1 and core proteins could be detected includes, but it is not limited to, serum, saliva and liver.

Of course, those skilled in the art would readily understand that the genotype-specific and universal peptides of the present invention and expression vectors containing nucleic acid sequence capable of directing host organism synthesis of these peptides could also be used as vaccines against hepatitis C. Formulations suitable for administering the peptide(s) and expression vectors of the present invention as immunogen, routes of administration,

372577\_1

o pharmaceutical compositions comprising the peptides expression vectors and so forth are the same as those previously described for recombinant El and core proteins.

The genotype-specific and universal peptides of the present invention and expression vectors containing nucleic acid sequence capable of direct host organism synthesis of these peptides may also be supplied in the form of a kit, alone, or in the form of a pharmaceutical composition as described above for recombinant El and core proteins.

Any articles or patents referenced herein are incorporated by reference. The following examples illustrate various aspects of the invention but are in no way intended to limit the scope thereof.

15

5

10

20

25

30

MATERIALS

Serum used in these examples was obtained from 84 anti-HCV positive individuals who were previously found to be positive for HCV RNA in a cDNA PCR assay with primer set a from the 5' NC region of the HCV genome (Bukh, J. et al. (1992 (b)) Proc. Natl. Acad. Sci. USA 89:4942-4946). These samples were from 12 countries: Denmark (DK); Dominican Republic (DR); Germany (D); Hong Kong (HK); India (IND); Sardinia, Italy (S); Peru (P); South Africa (SA); Sweden (SW); Taiwan (T); United States (US); and Zaire (Z).

10

15

20

25

30

35

5

### Example 1

Identification of the cDNA Sequence of the E1 Gene of 51 Isolates of HCV via RT-PCR Analysis of Viral RNA Using Universal Primers

Viral RNA was extracted from 100  $\mu$ l of serum by the quanidinium-phenol-chloroform method and the final RNA solution was divided into 10 equal aliquots and stored at -80°C as described (Bukh, et al. (1992 (a)). The sequences of the synthetic oligonucleotides used in the RT-PCR assay, deduced from the sequence of HCV strain H-77 (Ogata, N. et al. (1991) Proc. Natl. Acad. Sci. USA 88:3392-3396), are shown as SEO ID NOs: 207-212. One aliquot of the final RNA solution, equivalent to 10  $\mu l$  of serum, was used for cDNA synthesis that was performed in a 20  $\mu l$  reaction mixture using avian myeloblastosis virus reverse transcriptase (Promega, Madison, WI) and SEQ ID NO:208 as a primer. The resulting cDNA was amplified in a "nested" PCR assay by Tag DNA polymerase (Amplitag, Perkin-Elmer/Cetus) as described previously (Bukh et al. (1992a)) with primer set e (SEQ ID NOs:207-210). Precautions were taken to avoid contamination with exogenous HCV nucleic acid (Bukh et al. 1992a)), and negative controls (normal, uninfected serum) were interspersed between every test sample in both the RNA extraction and cDNA PCR procedures. No false positive results were observed in the analysis. In most instances,

amplified DNA (first or second PCR products) was reamplified with primers SEO ID NO:211 and SEQ ID NO:212 prior to sequencing since these two primers contained EcoR1 sites which would facilitate future cloning of the El gene. Amplified DNA was purified by gel electrophoresis followed by glass-milk extraction (Geneclean, BIO 101, LaJolla, CA) and both strands were sequenced directly by the dideoxynucleotide chain termination method (Bachman, B. et al. (1990) Nucl. Acids Res. 18:1309)) with phage T7 DNA polymerase (Sequenase, United States Biochemicals, Cleveland, OH), [alpha 35S] dATP (Amersham, Arlington 10 Heights, IL) or [alpha 33P] dATP (Amersham or DuPont, Wilmington, DE) and sequencing primers. RNA extracted from serum containing HCV strain H-77, previously sequenced by Ogata, N. et al. (1991), was amplified with primer set e (SEO ID NOs:207-210) and sequenced in parallel as a 15 control. The nucleotide sequences of the envelope 1 (E1) gene of all 51 HCV isolates are shown as SEQ ID NOs:1 - 51. In all 51 HCV isolates, the E1 gene was exactly 576

# Example 2

nucleotides in length and did not have any in-frame stop

Computer Analysis of the Nucleotide and Deduced Amino Acid Sequences of the El Gene of 51 HCV Isolates

Multiple computer-generated alignments of the nucleotide (SEQ ID NOs:1-51, Figures 1A-H) and deduced amino acid sequences (SEQ ID NOs:52-102, Figures 2A-H) of the cDNAs of the 51 HCV isolates constructed using the computer program GENALIGN (Miller, R.H. et al. (1990) Proc. Natl. Acad. Sci. USA 87:2057-2061) resulted in the 51 HCV isolates being divided into twelve genotypes based upon the degree of variation of the E1 gene sequence as shown in table 1.

20

25

30

codons.



- 64

Biochemistry: Bukh et al.

ii:	I/1a	II/1b	III/2a	IV/2b	3c	(V)/3a	4a	46	40	44	5a	<b>6a</b>		
6a	62.2-63.9	0.69-9.19	55.0-56.8		97.6	58.0-58.9	62.0	62.7	63.0-64.4	64.1	60.6-63.2	:		;
5a	62.3-67.2	62.2-66.5	8.09-9.95	53.5-56.6	56.9-57.1	61.8-64.4	62.8-64.6	63.9-64.6	62.7-64.8	64.4-66.1	90.1-95.7		92.7-97.4	62.0-63.5
4d	67.7-69.4 62.3-67.2	63.0-65.5	58.7-61.3	57.5-59.0	58.9	62.3-63.9	74.8	72.0	9.81-9.71			-	69.3-71.4	66.7
40	62.7-64.4	61.6-65.1	59.7-63.4	57.1-59.9 57.5-59.0 53.5-56.6 53.6-55.2	58.0-58.3	60.9-62.5	75.5-78.0	74.0-74.8	1.06		9.68	82.8	65.1-67.2 67.7-71.4	66.1-67.2
4b	64.9-66.8	63.4-65.8	58.9-60.4	56.4-57.6	58.5	62.7-64.1	74.8	1			79.2-80.2	77.6	65.1-67.2	62.5
4a	63.9-67.2	60.9-63.7	61.5-62.7	8.09-6.85	59.2	64.4-65.3			1	76.0	77.1-81.3	78.1	67.2-68.2	1.99
(V)/3a	63.0-66.3	63.9-67.2	58.0-60.8	56.3-58.3	57.5-58.2	93.8-99.1		94.3-98.4	66.1-68.8	62.0-64.6	63.0-65.6	63.5-64.6	60.4-64.1	57.8-58.9
20	60.8-62.8	60.1-61.5	72.7-73.6	67.5-68.9	1		i	52.1-53.6	58.3	53.6	54.7-58.3	54.2	54.2-56.8	50.5
IV/2b	56.1-58.3	53.8-57.5	69.1-71.0	92.7-95.0	L	93.8-96.4	8.69-1.19	54.2-56.8	58.9-60.4	52.1-53.1	54.2-58.3	55.2-55.7	50.5-53.1	49.0-50.5
III/2a	59.2-63.7	58.3-62.2	88.0-91.3		89.1-92.7	69.3-72.9	74.5-77.1	54.7-58.9	62.0-63.0	53.6-56.3	55.2-61.5	56.3-58.9	52.6-57.3	49.0-51.0
11/11	72.0-76.2	6.76-6.88		90.1-97.9	52.6-56.8	51.0-54.2	52.6-55.7	8.07-7.99	64.6-67.2	66.1-70.3	64.6-69.3	66.7-70.3	64.1-70.3	62.5-65.6 49.0-51.0
I/Ia	9.76-6.68		91.1-98.4	75.5-80.7	58.3-64.6	54.2-56.8	56.3-60.4	64.1-68.8	69.3-73.4	66.7-69.3	66.1-72.9	73.4-75.5	66.1-73.4	64.6-65.6
		aa:	Vla	11/19	III/2a	IV/2b	20	(V)/3a	4a	4	4	P4	5a	е9

Nucleotide sequences analyzed in compiling the above table are shown in SEQ ID NOs:1-51 while the amino acid sequences analyzed are shown in SEQ ID NOs:52-102. The grouping of SEQ ID NOs: into genotypes is previously described in the specification.

15

20

25

30

35

The nucleotide and amino acid sequence identity of HCV isolates of the same genotype was in the range of 88.0-99.1% and 89.1-98.4%, respectively, whereas that of HCV isolates of different genotypes was in the range of 53.5-78.6% and 49.0-82.8%, respectively. The latter differences are similar to those found when comparing the envelope gene sequences of the various serotypes of the related flaviviruses, as well as other RNA viruses. When microheterogeneity in a sequence was observed, defined as more than one prominent nucleotide at a specific position, the nucleotide that was identical to that of the HCV prototype (HCV1, Choo et al. (1989)) was reported if possible. Alternatively, the nucleotide that was identical to the most closely related isolate is shown.

Analysis of the consensus sequence of the E1 protein of the 51 HCV isolates from this study demonstrated that a total of 60 (30.3%) of the 192 amino acids of the E1 protein were invariant among these isolates (Fig. 3). impressive, all 8 cysteine residues as well as 6 of 8 proline residues were invariant. The most abundant amino acids (e.g. alanine, valine and leucine) showed a very low degree of conservation. The consensus sequence of the E1 protein contained 5 potential N-linked glycosylation sites. Three sites at positions 209, 305 and 325 were maintained in all 51 HCV isolates. A site at position 196 was maintained in all isolates except the sole isolate of genotype 2c. Also, a site at position 234 was maintained in all isolates except one isolate of genotype I/la, all four isolates of genotype IV/2b and the sole isolate of genotype 6a. Conversely, only genotype IV/2b isolates had a potential glycosylation site at position 233. Further analysis revealed a highly conserved amino acid domain (aa 302-328) in the E1 protein with 20 (74.1%) of 27 amino acids invariant among all 51 HCV isolates. It is possible that the 5' and 3' ends of this domain are conserved due to important cysteine residues and N-linked glycosylation

10

15

sites. The central sequence, 5'-GHRMAWDMM-3' (aa 315-323), may be conserved due to additional functional constraints on the protein structure. Finally, although the amino acid sequence surrounding the putative El protein cleavage site was variable, an amino acid doublet (GV) at position 380 was invariant among all HCV isolates.

A dendrogram of the genetic relatedness of the E1 protein of selected HCV isolates representing the 12 genotypes is shown in Fig. 4. This dendrogram was constructed using the program CLUSTAL (Higgins, D.G. et al. (1988) Gene, 73:237-244) and had a limit of 25 sequences. The scale showing percent identity was added based upon manual calculation. From the 51 HCV isolates for which the complete sequence of the E1 gene region was obtained, 25 isolates representing the twelve genotypes were selected for analysis. This dendrogram in combination with the analysis of the E1 gene sequence of 51 HCV isolates in Table 1 demonstrates extensive heterogeneity of this important gene.

The worldwide distribution of the 12 genotypes among 74 HCV isolates is depicted in Fig. 5. The complete 20 E1 gene sequence was determined in 51 of these HCV isolates (SEQ ID NOs:1-51), including 8 isolates of genotype I/1a, 17 isolates of genotype II/1b and 26 isolates comprising genotypes III/2a, IV/2b, 2c, 3a, 4a-4d, 5a and 6a. remaining 23 isolates, all of genotypes I/la and II/lb, the 25 genotype assignment was based on a partial El gene sequence since they did not represent additional genotypes in any of the 12 countries. The number of isolates of a particular genotype is given in each of the 12 countries studied. Of the twelve genotypes, genotypes I/la and II/lb were the 30 most common accounting for 48 (65%) of the 74 isolates. Analysis of the E1 gene sequences available in the GenBank data base at the time of this study revealed that all 44 such sequences were of genotypes I/la, II/lb, III/2a and IV/2b. Thus, based upon El gene analysis, 8 new genotypes 35

10

15

20

25

30

35

of HCV have been identified.

Also of interest, different HCV genotypes were frequently found in the same country, with the highest number of genotypes (five) being detected in Denmark. Of the twelve genotypes, genotypes I/la, II/lb, III/2a, IV/2b and V/3a were widely distributed with genotype II/lb being identified in 11 of 12 countries studied (Zaire was the only exception). In addition, while genotypes I/la and II/lb were predominant in the Americas, Europe and Asia, several new genotypes were predominant in Africa.

It was also found that genotypes I/la, II/lb, III/2a, IV/2b and V/3a of HCV were widely distributed around the world, whereas genotypes 2c, 4a, 4b, 4d, 5a and 6a were identified only in discreet geographical regions. For example, the majority of isolates in South Africa comprised a new genotype (5a) and all isolates in Zaire comprised 3 new closely related genotypes (4a, 4b, 4c). These genotypes were not identified outside Africa.

## Example 3

# Identification of the cDNA Sequence Of The Core Gene Of 52 Isolates Of HCV

Viral RNA extraction, cDNA synthesis and "nested" PCR were carried out as in Example 1. For the cDNA PCR assay HCV-specific synthetic oligonucleotides deduced from previously determined sequences that flank the C gene were used. Amplified DNA was purified by gel electrophoresis followed by glass-milk extraction as described in Example 1 or by electroelution and both strands were sequenced directly. In 44 of the 52 HCV isolates studied the procedures for direct sequencing described in Example 1 were utilized. For a number of the HCV isolates confirmatory sequencing was performed with the Applied Biosystems 373A automated DNA sequencer and 8 HCV isolates of genotype I/la or II/lb were sequenced exclusively by this method. All 73 negative control samples interspersed

10

15

20

25

30

35

among the test samples were negative for HCV RNA.

The amplified DNA fragment obtained in 50 of the 52 HCV isolates was specifically designed to overlap with previously obtained 5'NC sequences (Bukh et al. (1992b) Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946) and with the E1 sequences disclosed herein at approximately 80 nucleotide positions each. A complete match was observed in 6033 of 6035 overlapping nucleotides. Two discrepancies were observed in isolate US6 at nt 552 (C and T) and nt 561 (C and T) respectively. This may have been due to microheterogeneity at these nucleotide positions, since the remaining overlapping sequence was unique for isolate US6. In addition, there were 3 confirmed instances of microheterogeneity: nt 33 in isolate SA11 (C,T and T), nt 36 in isolate S45 (A,C and A), and nt 552 in isolate P10 (C,T and T). Overall, the excellent agreement in these overlapping sequences in this study with the NC sequences disclosed in Bukh et al. and with the El sequences disclosed herein definitively ruled out contamination as a source of non-authentic HCV sequences. Furthermore, this analysis proved that the sequences obtained were from a single population, and not from different populations as could happen in mixed infections.

The core (C) gene was exactly 573 nucleotides in length in all 52 HCV isolates with an amino terminal start codon and no in-frame stop codons. Microheterogeneity was observed in 26 of the 52 HCV isolates at 0.2-1.4% of the 573 nucleotide positions of the C gene, and resulted in changes in 0.5-1.0% of the 191 predicted amino acids in 12 of these isolates. A multiple sequence alignment was performed and it showed that the nucleotide identities of the C gene among these HCV isolates were in the range of 79.4-99.0%. In order to compare the genetic relatedness of HCV isolates in different gene regions, phylogenetic trees of the C gene of all 52 HCV isolates and the E1 gene of 51 HCV isolates were constructed using the unweighted pair-

group method with arithmetic mean (Nei, M. (1987) Molecular Evolutionary Genetics (Columbia University Press, New York, N.Y., pp. 287-326) (Figure 8). In both dendrograms a division of the 45 HCV isolates from which C and El genes had been cloned into at least six major genetic groups 5 (genotypes 1-6) and 12 minor genetic groups (genotypes I/la, II/lb, III/2a, IV/2b, 2c, V/3a, 4a-4d, 5a, and 6a) was observed. It is noteworthy that a major division in genetic distance between HCV isolates of genotype 2 and those of the other genotypes in the phylogenetic analyses of both gene sequences was observed. Furthermore, the 10 divergence of the minor genotypes within genotype 2 exhibited a degree of heterogeneity that is equivalent to that observed among the major genotypes. Analysis of the C gene from isolates Z5 and Z8, which had a unique 5' NC sequence (Bukh et al. (1992)) but from which the E1 gene 15 could not be amplified, revealed that these isolates represented two additional genotypes. The designations 4e and 4f are assigned to these genotypes that have not been described previously. Overall, the present specification demonstrates that the genetic relatedness of HCV isolates 20 is equivalent when analyzing the most conserved gene (C) and one of the most variable genes (E1) of the HCV genome, thereby providing strong evidence for the suggested division into major and minor genotypes.

25

30

35

#### Example 4

Computer Analysis of the Nucleotide and Deduced Amino Acid Sequences Of The Core Gene Of 52 HCV Isolates

In order to study further the heterogeneity of the C gene, a consensus sequence of the core gene from the 52 HCV isolates (Fig. 6J) was obtained. A total of 335 (58.5%) of the 573 nucleotides of the C gene were invariant among these HCV isolates. Nucleotides at the 1st and 2nd codon positions were invariant at 70.7% and 81.7% of these positions, respectively, while nucleotides at the 3rd

position were invariant at only 23.0% of such positions. Stretches of 6 or more invariant nucleotides were observed from nucleotides 1-8, 22-27, 85-92, 110-125, 131-141, 334-340, 364-371, 397-404, and 511-516 and may be suitable for anchoring primers for amplification of HCV RNA in cDNA PCR sasavs.

Genotype-specific nucleotide positions of the core gene of hepatitis C virus were also noted for each of the genotypes. These genotype-specific nucleotides are shown below where each genotype-specific nucleotide is given in parentheses next to the nucleotide position in which it is found.

<u>Genotype 1:</u> 460 (C), 466 (C), 483 (C), 486 (G).

Genotype I/la: 180 (T).

Genotype II/1b: 106 (C), 273 (G).

<u>Genotype 2:</u> 192 (C), 201 (A), 203 (A), 207 (G), 210 (C), 221 (A), 231 (A), 232 (A), 341(A).

20

10

15

Genotype III/2a: 315 (C), 355 (G).

<u>Genotype IV/2b:</u> 45 (A), 174 (G), 216 (C), 348 (A), 376 (A), 414 (T).

25

<u>Genotype 2c:</u> 233 (G), 312 (C), 318 (A), 456 (C), 462 (G), 543 (C), 556 (T).

Genotype V/3a: 47 (T), 84 (A), 106 (G), 126 (A), 150 (T), 30 212 (G), 216 (A), 300 (A), 491 (T), 559 (C), 560 (A), 568 (G), 571 (A), 572 (G)

Genotype 4: 59 (T).

35 <u>Genotype 4a:</u> 213 (A), 231 (G), 415 (A).

15

20

25

30

35

Genotype 4b: 66 (G), 145 (G), 310 (A).

Genotype 4c: 213 (T), 219 (A), 270 (T).

Genotype 4d: 212 (T), 327 (G), 469 (C).

Genotype 4e: 199 (C), 306 (A), 326 (A).

Genotype 4f: 57 (T), 75 (A), 267 (A).

10 Genotype 5a: 291 (G), 294 (C).

Genotype 6a: 59 (C), 175 (A), 195 (A), 198 (A), 214 (C), 224 (A), 316 (C), 351 (G), 387 (G), 444-447 (GGCT), 450 (G), 471-472 (AA), 474 (C).

These genotype-specific nucleotides are of utility in designing the genotype-specific PCR primers and hybridization probes.

Finally, although the full length nucleic acid sequence of the C gene of isolates representing genotypes I/la, III/lb, IIII/2a, IV/2b and V/3a have been reported by others, those of 9 of the 14 genotypes (i.e., 2c, 4a-4f, 5a and 6a) have not been reported previously. In sum, by aligning the consensus sequences of the major genotypes, the present application enables those skilled in the art to map universally conserved sequences as well as genotypespecific sequences of the C gene among 14 genotypes of HCV.

In order to study the heterogeneity of the deduced C protein, a multiple sequence alignment of the predicted amino acids for all 52 HCV isolates was performed, and a consensus sequence was obtained (Fig. 7J). The identities of the predicted 191 amino acids of the C protein among these HCV isolates were in the range of 85.3-100.0%. A total of 132 (69.1%) of the 191 amino acids of the C protein were invariant. The most prevalent amino acids in the consensus sequence were glycine (13.6%),

arginine (12.6%), proline (11.0%), and leucine (9.9%). The most conserved amino acids were tryptophan (5 of 5 amino acids invariant), aspartic acid (5 of 5 amino acids invariant), proline (19 of 21 amino acids invariant) and glycine (23 of 26 amino acids invariant). Previous
analyses indicated that HCV is evolutionarily related to pestiviruses (Miller et al. (1990) Proc. Natl. Acad. Sci.

analyses indicated that HCV is evolutionarily related to pestiviruses (Miller et al. (1990) Proc. Natl. Acad. Sci. U.S.A. 87:2057-2061). In this regard, it is of interest to note that the C proteins of both viruses have a high content of proline residues (Collette M.S. et al. (1988) Virology 165:200-208) which are likely to be important in

10 <u>Virology</u> 165:200-208), which are likely to be important in maintaining the structure of this protein. As is characteristic for a protein that binds to nucleic acid, the C protein has conserved amino acids that are basic and positively charged, and these are capable of neutralizing the negative charge of the HCV RNA encapsidated by this protein (Rice, C.M. et al. (1986) in Togaviridae and Flaviviridae, eds Schleinger, S. & Schlensinger, M.J. (Plenum Press, New York, N.Y.) pp. 279-326). Specifically,

over 16% of the amino acids in the consensus sequence of
the C protein of HCV are arginine and lysine that are
located primarily in three clusters (i.e., from amino
acids 6-23, 39-74 and 101-121) (Shih, C.M. et al. (1993) J.
Gen. Virol. 67:5823-5832) (Fig. 7J). The 10 arginine and
lysine residues within amino acids 39-62 are invariant
among all 52 HCV isolates, suggesting that this domain may

among all 52 HCV isolates, suggesting that this domain may represent an important RNA-binding site. The capsid proteins of the related flavi-and pestiviruses (Miller et al. (1990)) also have a high content of arginine and lysine (Rice et al. (1986); Collette et al. (1988). Although there are three major hydrophilic regions (i.e., amino

there are three major hydrophilic regions (i.e., amino acids 2-23, 39-74 and 101-121) that are conserved in all 52 HCV isolates, the remainder of the C protein is hydrophobic. Interestingly, one such highly conserved hydrophobic domain from aa 24-39 is flanked by proline residues. The hydrophobic domains are likely to be

20

25

involved in protein-protein and/or protein-RNA interactions during assembly of the nucleocapsid, as well as in interaction with the lipoprotein envelope, as has been suggested for flaviviruses (Rice et al. (1986)). significant observations are: (i) a cluster of 5 invariant 5 tryptophan residues from aa 76-107; (ii) the lack of an Nlinked glycosylation site (N-X-T/S); (iii) two potential nuclear localization signals (i.e., PRRGPR at amino acids 38-43 and PRGRRQP at amino acids 58-64) that are present in all 52 HCV isolates (Shih et al. (1993)); and (iv) a putative DNA-binding motif SPRG at amino acids 99-102, 10 found in 51 of the 52 HCV isolates, with SP present in all 52 isolates. This study demonstrates that the C protein has features that are highly conserved among the various genotypes of HCV, and that are known to be characteristic

of capsid proteins of other related viruses.

It should also be noted that the phylogenetic analysis of the amino acid sequence of the C proteins was not capable of resolving the minor groups within genotypes 1 and 4 because of the conservation of this protein (data not shown). Indeed, only a few type-specific amino acids were identified. One striking example was that isolates of genotype 4 have an additional methionine at position 20 that is specific for this major genetic group. Finally, the conservation of the sequences surrounding the cleavage site between the C and the E1 proteins of the different genotypes, which has been determined to be between amino acid 191 (alanine) and aa 192 (tyrosine) in HCV isolates of genotype 1 was analyzed (Hijikata, M., et al. (1991) Proc. Natl. Acad. Sci. USA 88:5547-5551). The C-terminal

sequence of C is serine-alanine in all but one of the 48 HCV isolates comprising genotypes 1, 2, 4, 5 and 6. However, all 4 HCV isolates of genotype 3 in this study, as well as isolates of genotype 3 published previously (Okamoto, H., et al. (1993) J. Gen. Virol. 74:2385-2390, Stuyver, L., et al. (1993) Biochem. Biophys. Res. Comm.

15

20

25

30

35

192:635-641), contain alanine-serine at this position. Thus, studies will be needed to determine the C/E1 cleavage site in genotype 3 isolates. Overall, the present invention application discloses the mapping of universally conserved sequences, as well as genotype-specific
5 sequences, of the C protein among 14 genotypes of HCV.

Implications of the mapping of universally conserved and genotype-specific core nucleotide and amino acid core sequences for diagnosis of HCV infection and for determination of HCV genotypes

Detection of antibodies directed against the HCV core protein is important in the diagnosis of HCV infection. The recombinant C22-3 protein, spanning amino acids 2-120 of the C gene, is a major component of the commercially available second-generation anti-HCV tests. Several studies have indicated that the three major hydrophilic regions of the C protein contain linear immunogenic epitopes (summarized in <u>J. Clin. Microbiol</u>, 30:1989-1994) (Sällberg, M. et al. (1992). For example, antibodies against synthetic peptides from amino acids 1-18, 51-68 and 101-118 were detected in infected patients (Sällberg, M. et al. (1992)). The present application demonstrates that, while these immunogenic regions are

highly conserved, genotype-specific differences are observed at several amino acid positions that may influence the specificity and sensitivity of the serological tests. One such example is that a single amino acid substitution at amino acid 110 has been demonstrated to affect sero-reactivity (Sällberg, et al. (1992)). Despite the high degree of conservation in the immunodominant regions of the C protein among the different genotypes, it is possible

C protein among the different genotypes, it is possible that genetic heterogeneity of the C protein could lead to false negative results in current serological tests.

With respect to genotype analysis, several methods have been used to determine the genotype of HCV isolates without resorting to sequence analysis. These

include PCR followed by: (i) amplification with typespecific primers (Okamoto, H. et al. (1992) J. Gen. Virol., 73:673-679); (ii) determination of restriction-length polymorphism (Simmons, P. et al. (1993) J. Gen. Virol., 74:661-668); and (iii) specific hybridization (Stuyver, L. 5 (1993) J. Gen. Virol., 74:1093-1102). The proposed methods have primarily been based on 5' NC and C sequences. Previous studies suggested that 5' NC-based genotyping systems would only be predictive of the major genetic groups of HCV (Bukh, J., et al. (1992) Proc. Natl. Acad. 10 Sci. USA 89:4942-4946, Bukh, J., et al. (1993) Proc. Natl. Acad. Sci. USA 90:8234-8238). The most widely used Cbased genotype system has been the PCR assay with typespecific primers that was designed for distinguishing HCV isolates of genotypes I/la, II/lb, III/2a, IV/2b and V/3a

(Okamoto, H., et al. (1993) J. Gen. Virol. 74:2385-2390,
Okamoto, H. et al. (1992) J. Gen. Virol. 73:673-679).
Since this system was developed prior to the identification
of genotypes 2c, 4a-4f, 5a and 6a there are significant
limitations to this typing system. For example, the
primers specific for genotype IV/2b (nt 270-251) are as

20 primers specific for genotype IV/2b (nt 270-251) are as highly conserved within isolates of genotype 4c and 6a as within the isolates of genotype IV/2b. Thus, this assay probably can not distinguish among these genotypes. Another C-based approach involves distinguishing between genotypes 1 and 2 by type-specific antibody responses (Machida et al (1992) Hepotology, 16:886-891). Synthetic peptides

(1992) <u>Hepotology</u>, 16:886-891). Synthetic peptides composed of amino acids 65-81 were found to be genotypespecific for genotypes 1 and 2 in ELISA assays. The present analysis of amino acid sequences demonstrated significant variation within isolates of genotypes 1 and 2.

30 significant variation within isolates of genotypes 1 and 2. Thus it is likely that these peptides will not identify all isolates of genotypes 1 and 2. Furthermore, the peptide for genotype 1 was highly conserved within isolates of genotypes 3 and 4 and might detect antibodies against these genotypes as well. Finally, it should be pointed out that

10

15

20

25

30

35

most isolates of genotypes 3 and 4 had an identical amino acid sequence at positions 65-81.

## Example 5

# Detection by ELISA Based on Antigen from Insect Cells Expressing Complete E1 Or Core Protein

Expression of E1 or Core protein in SF9 cells. A cDNA (eg SEQ ID NO:1) encoding a complete E1 protein (eg SEQ ID NO:52) or a cDNA (eg SEQ ID NO:103) encoding a complete core protein (e.g., SEQ ID NO:155) is subcloned into pBlueBac - Transfer vector (Invitrogen) using standard subcloning procedures. The resultant recombinant expression vector is cotransfected into SF9 insect cells (Invitrogen) by the Ca precipitation method according to the Invitrogen protocol.

ELISA Based on Infected SF9 cells.  $5 \times 10^{6} \text{ SF}$ cells infected with the above-described recombinant expression vector are resuspended in 1 ml of 10 mM Tris-HCl, pH 7.5, 0.15M NaCl and are then frozen and thawed 3 times. 10 ul of this suspension is dissolved in 10 ml of carbonate buffer (pH 9.6) and used to cover one flexible microtiter assay plate (Falcon). Serum samples are diluted 1:20, 1:400 and 1:8000, or 1:100, 1:1000 and 1:10000. Blocking and washing solutions for use in the ELISA assay are PBS containing 10% fetal calf serum and 0.5% gelatin (blocking solution) and PBS with 0.05% Tween -20 (Sigma, St.Louis, MO) (washing solution). As a secondary antibody. peroxidase-conjugated goat IgG fraction to human IgG or horse radish peroxidase-labelled goat anti-Old or anti-New World monkey immunoglobulin is used. The results are determined by measuring the optical density (O.D.) at 405 nm.

To determine if insect cells-derived E1 or core protein representing genotype  ${\rm I/a}$  of HCV could detect anti-HCV antibody in chimpanzees infected with genotype  ${\rm I/la}$  of HCV, three infected chimpanzees are examined. The serum of

all 3 chimpanzees are found to seroconvert to anti-HCV.

## Example 6

## Use of the Complete El Protein as a Vaccine

Mammals are immunized with purified or partially purified El protein in an amount sufficient to stimulate the production of protective antibodies. The immunized mammals challenged with various genotypes of HCV are protected.

It is understood by one skilled in the art that the recombinant El protein used in the above vaccine can also be used in combination with other recombinant El proteins having an amino acid sequence shown in SEQ ID NOs:52-102. In addition, recombinant core proteins having an amino acid sequence shown in SEQ ID NOs:155-206 could also be used in the above vaccine, either alone, in combination with other recombinant core proteins of the present invention, or in combination with recombinant El proteins having an amino acid sequence shown in SEQ ID NOs:52-102.

## Example 7

Determination of the Genotype of an HCV Isolate Via Hybridization of Genotype-Specific Oligonucleotides to RT-PCR Amplification Products.

Viral RNA is isolated from serum obtained from a mammal and is subjected to RT-PCR as in Example 1 or Example 3. Following amplification, the amplified DNA is purified as described in Example 1 or Example 3 and aliquots of 100 ul of amplification product are applied to dots on a nitrocellulose filter set in a dot blot apparatus. The dots are then cut into separate dots and each dot is hybridized to a <sup>32</sup>P-labelled oligonucleotide specific for a single genotype of HCV. The oligonucleotides to be used as hybridization probes are

372577 1

35

5

10

15

20

2.5

deduced from the consensus sequences shown in Figures 1A-1H or 6A-6J or from the SEQ ID NOs: representing E1 or core sequences comprising genotypes 4a-4f, 2c and 6a.

## Example 8

# ELISA Based on Synthetic Peptides Derived From E1 cDNA Sequences

E1 peptide(s) specific for genotype I/la is placed in 0.1% PBS buffer and 50ul of a lmg/ml solution of peptide is used to cover each well of the microtiter assay plate. Serum samples from two mammals infected with genotype I/la HCV and from one mammal infected with genotype 5a HCV are diluted as in Example 3 and the ELISA is carried out as in Example 3. Both mammals infected with genotype I HCV react positively with peptides while the mammal infected with genotype 5a HCV exhibits no reactivity. One skilled in the art would readily understand that in the above experiment, core peptides specific for genotype I/la could be used in place of, or in combination with the El genotype-specific peptide(s).

# Example 9

# Use of El Peptides as a Vaccine

Since the E1 genotype-specific peptides of the present invention are derived from two variable regions in the complete E1 protein, there exists support for the use of these peptides as a vaccine to protect against a variety of HCV genotypes. Mammals are immunized with peptide(s) selected from SEQ ID NOs: 136-159 in an amount sufficient to stimulate production of protective antibodies. The immunized mammals challenged with various genotypes of HCV are protected. One skilled in the art would readily understand that genotype-specific core peptides of the present invention could also be used either alone, in combination with each other, or in combination with the

35

30

5

10

15

20

genotype-specific El peptides, as a vaccine to protect against a variety of HCV genotypes. In addition, the above vaccines may also be formulated using the universal core and/or El peptides of the present invention.

5

10

15

20

25

30

# SEQUENCE LISTING

	(1)	GENERAL :	INFORMATION:
		(i)	APPLICANTS: BUKH, J., MILLER, R.H. AND PURCELL, R.H.
5		(ii)	TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
10		(iii)	NUMBER OF SEQUENCES: 263
15		(iv)	CORRESPONDENCE ADDRESS: (A) ADDRESSEE: MORGAN & FINNEGAN, L.L.P. (B) STREET: 345 PARK AVENUE (C) CITY: NEW YORK (D) STATE: NEW YORK (E) COUNTRY: USA (F) ZIP: 10154
		(v)	COMPUTER READABLE FORM: (A) MEDIUM TYPE: FLOPPY DISK (B) COMPUTER: IBM PC COMPATIBLE (C) OPERATING SYSTEM: PC-DOS/MS-DOS (D) SOFTWARE: WORDPERFECT 5.1
20		(vi)	CURRENT APPLICATION DATA: (A) APPLICATION NUMBER: TO BE ASSIGNED (B) FILING DATE: 26-MAY-1998
25		(vii)	PRIOR APPLICATION DATA: (A) APPLICATION NUMBER: 08/290,665 (B) FILING DATE: 15-AUG-1994
		(viii)	CURRENT APPLICATION DATA: (A) APPLICATION NUMBER: 08/086,428 (B) FILING DATE: 29-JUNE-1993
30		(ix)	ATTORNEY/AGENT INFORMATION: (A) NAME: RICHARD W. BORK (B) REGISTRATION NUMBER: 36,459 (C) REFERENCE/DOCKET NUMBER: 2026-4116US2
		(x)	TELECOMMUNICATION INFORMATION: (A) TELEPHONE: (212) 758-4800 (B) TELEFAX: (212) 751-6849 (C) TELEX: 421792

0		
	(2) INFORMATION FOR SEQ ID NO:1:	
5	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: DK7	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:	
10	TAC CAA GTG CGC AAC TCC ACG GGG CTT TAC CAT GTC ACC AAT GAT TGC CCT AAC TCG AGT ATC GTG TAC GAG GCG GCC GAT GCC ATC CTG CAC ACT CCG GGG TGT GTC CCT TGC GTT CGC GAG GGT AAC GTC TCG AGG TGT TGG GTG GCG ATG ACC CCC ACG GTG GCC ACC AGG GAT GCC ACC CTC TGC GTG CAC GTC CAC GGG CAT CTC CTC GGG AGT GCC ACC CTC TGT TCG GCC CTC TAC GTG GGG GAC CTT TGC GGG	3: 7: 11' 15: 19: 23:
15	TCT GTC TTT CTT GTC GGT CAA CTG TTT ACC TTC TCT CCC AGG CGC CAC TGG ACG ACG CAA GGC TGC AAT TGT TCT ATC TAT CCT GGC CAT ATA ACG GGT CAC CGC ATG GGC TGG GAT ATG ATG ATG ATG ACT GGT CAC ACG GCG TTG GTA GTA GCT CAG CTG CTC CGG ATC CCG CAA GCC ATC TTG GAC ATG ATC GCT GGT GGT CAC TGG GGA GTC CTG GGC GGC ATA GCG TAT TTT TCC ATG GTG GGG AAC TGG GCC AAC GTG CTG GTA GTG GTG CTG CTG CTG TTG GTA GTG GTG CTG CTG CTG CTG GTA GTG CTG CTG CTG CTA TTT GCC GGC GTC GAC GCG	312 353 390 429
20	and and an iii add add are are acc	576
	(2) INFORMATION FOR SEQ ID NO:2:	
25	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: DK9	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:	
30	TAC CAA GTA CGC AAC TCC TCG GGC CTC TAC CAT GTC ACC AAT GAT GAT GCC CAT AAC TCG AGT ATT GTG TAC GAG GCG GCC GAT GCC ATC CTG CAT TCT CCA GGG TGT GTC CCT TGC GTT CCC GAG GGT AAC GCC TCG AAA TGT TGG GTG GCC GCG ACG GCC ACG GTG GCC ACC ACG CTT CGA CTC CTC GAT TCT CTC GAG GAC CTC CCC GCA ACG CAC CTC TGC TCG CTC TAT GTG GGG GAC TTTG TGC GGG	39 78 113 156 195 234
35		2/3

۰		
5	TCT GTC TTC CTT GTC GGC CAA CTG TTC ACC TTC CCC CCC AGA CGC CAC TGG ACA ACG CAA GAC TGC AAC TGC ATC TCT ATC TAC CCC GGC CAT ATT ACG GGT CAT CGC ATG GGC TGG GAT ATG ATG ATG ATG ATG ATG CCC GCC CAC ACC CAC CAC CAC CTG GTA ATG GCC CGC CTG CTA ACG CAC CTG CTA ACG CAC CTG CTA ACG CAC CTG CTC ACG ATC CCC CAC GCC ATC TTG GAC ATG ATC GCT GGT GCC CAC TGG GGA ACC CTA CCC GCC ATA GCC ATA CCC TAC TTC TCC ATG GTG GGC AAC TGG GCC AAC GCC ATA GCG CTA CTG TTC TCC ATG CTG GGC AAC TCC GTG GTG ACC CTG CTG CTG CTG CTG CTG CTG CTG CTG C	312 351 390 429 468 507 546 576
	(2) INFORMATION FOR SEQ ID NO:3:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: DR1	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:	
20	CAC CAA GTG CGC AAC TCT ACA GGG CTT TAC CAT GTC ACC AAT GAT TGC CCT AAT TCG AGT ATT GTG TAC GAG GGC GCC GAT GGC ATC CTG CAC GGC CGC GGG GTG TCC CCT GAC GGC CGC GGG GTG TCC CCT GAC GCC CCC ACG GTG ACC CCC ACG GTG ACC ACC ACC ACC GTT CAC CCC ACA ACC CCC ACC GTT CAC ACC CCC ACC GTT CAC ACC CCC CCC ACC GTT CAC ACC CCC ACC GTT CAC ACC CCC ACC ACC CTC TAC GAC CTC TTC TCT CGC GGG GC CAC CTC TCC CTC TCC CCC ACC ACC CCC ACC CCC C	39 78 117 156 195 234 273 312 351 390 429 468
25	ATC GCT GGA GCC CAC TGG GGA GTC CTA GCG GGC ATA GCG TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC GTA GTG CTG TTG CTG TTT GCC GGC GTT GAT GCG	507 546 576
30	(2) INFORMATION FOR SEQ ID NO:4:  (i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE:	5/6
	(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: DR4	

	(xi)	SEQUENCE	DESCRIPTION: SEQ ID NO:4:
5	GAT GCC ATC CGC GAG GGT CCC ACG GTG CAG CTC CGA ACC CTC TGC TCT GTC TTC AGG CAC CAC TAT CCC GGC	CTG CAC ACG AAC ACC TCG GCC ACC AGG CGT CAC ATC TCG GCC CTC CTT GTC GGT TGG ACA ACG CAT ATA ACG	ACA GGG CTT TAC CAT GTC ACC AGT ATT GTG TAC GAG GCG GCC AGT ATT GTG TAC GAG GCG GCC AGT GTG GTG GTG GCG GTG ACC AGA CGC AAA CTC CCC ACA ACG ACG GGG AAA CTC CCC ACA ACG ACG GTG GTT GTC GGG ACC ACG GTG GTT GTC GGG ACC ACA CTG CTT GTC GGG ACC ACA CTG CTT GTC GGG ACC ACA CTG CTT GTC TCC ACA GAC TCC ACT TCT CTC CCC ACA GAC TCC ACT TCT CTC CCC ACA GAC TCC ACT TCT CCC ACA GAC TCC ACT TCT CCC ACA GAC TCC ACT GCG ACC
10		CTC CGG ATC GCC CAC TGG ATG GTG GGG CTG TTT GCC	CCT ACG ACA GCG CTG GTA GTA  CCA CAA GCC ATC TTG GAC ATG 468 GGA GTC CTA GCG GGC ATA GCG AAC TGG GCG AAG GTC CTG GTA  GGC GTT GAT GCG  OLD NO.5.
15	(i)	SEQUENCE (A) LEN (B) TYP (C) STR	CHARACTERISTICS: GTH: 576 base pairs E: nucleic acid ANDEDNESS: single OLOGY: linear
20	(vi)	(C) IND	SOURCE: ANISM: homosapiens IVIDUAL ISOLATE: S14
25	CGT GAG GGT CCC ACG GTG CAG CTT CGA ACC CTC TGT TCT GTC TTT	AAC ACC TCG GCC ACC AGG CGT TAC ATC TCG GCC CTC CTT GTC GGT	ACG GGG CTT TAC CAT GTT ACC AGT ATT GTG TAC GAG ACA GCT 78 CCG GGA TGT GTC CCT TGC GTT 117 AGG TGT TGG GTG GCG ATG ACC 156 GAC GGC AAA CTC CCC GCA ACG 195 GAT CTG CTT GTC GGG AGC GCC 234 CTAC GTG GGG GAC TTG TGC GGG 273 CAG CTG TTT ACC TTC TCT CCC 312
30	TAT CCC GGC ATG ATG ATG GCT CAG CTG ATC GCT GGT TAT TTC TCC	CAT ATA ACG AAC TGG TCC CTC CGG ATC GCT CAC TGG ATG GTG GGA	G CAA GAC TGC AAT TGT TCT ATC  GGT CAT CGC ATG GCA TGG GAT  90 CCT ACG ACG GCA CTG GTA GTA  CCC ACG ACG GCA CTT GTA GTA  429  CCA CAA GCC ATC TTG GAT ATG  468  GGA GTC CTA GCG GGC ATA GCG  507  AAC TGG GCG AAG GTC CTA GTG  546  GGC GTT GAC GCG  576

35

(2) INFORMATION FOR SEQ ID NO:6:

	(1) SEQU (A) (B) (C) (D)		
5	(vi) ORIG (A) (C)	INAL SOURCE: ORGANISM: homosapiens INDIVIDUAL ISOLATE: S18	
	(xi) SEQU	ENCE DESCRIPTION: SEQ ID NO:6:	
10	AAT GAC TGC CCT AAC GAT ACC ATC CTA CAC CGC GAG GGT AAC GCC	TCC ACG GGC CTT TAC CAT GTC ACC TCG AGC ATT GTG TAC GAG ACG GCC TCT CCG GGG TGT GTC CCT TGC GTT TCG AGA TGT TGG GTG CCG GTG GCC AGG GAC GGC AAA CTC CCC GCA ACG	39 78 117 156 195
15	CAG CTT CGA CGT CAC ACC CTC TGC TCG GCC TCT GTC TTT CTT GTC AGG CGC CAC TGG ACA ATA CC GGC CAT ATA ATG ATG ATG AAC TGG GCT CAG CTG CTC AGG ATT CAG GCT GGT CAC CAC	ATC GAT CTG CTT GTT GGG AGC GCC CTC TAT GTG GGG AGC GCC ACG CAG CTG TTC ACT ATC TCC ACG CAA GAC TGC AAC TGT TCT ATC ACG GGT CAC CGT ATG GCA TGG AAT CCC CCT ACA AGC GGG TTG GTA ATA GTC CCC CAA GCC GTC TTG GAC ATG TGG GGA GTC CTA GCG GGC ATA GCG GGG AAC TGG GCA AAG GTC CTG CTA	234 273 312 351 390 429 468 507
	GTG CTG TTG CTG TTT	GCC GGC GTC GAT GCG	576
20	(2) INFORMATION FO	R SEQ ID NO:7:	
25	(i) SEQU (A) (B) (C) (D)		
23		INAL SOURCE: ORGANISM: homosapiens INDIVIDUAL ISOLATE: SW1	
	(xi) SEQU	ENCE DESCRIPTION: SEQ ID NO:7:	
30	AAT GAT TGC CCT AAC GAT GCC ATT CTA CAC CGC GAG GAT GGC GCC CCC ACA GTC GCC ACT CAG CTT CGA CGT CAC ACC CTC TGC TCG GCC	TCC TCG GGC CTT TAC CAT GTC ACC TCG AGT ATT GTG TAC GAG ACG GCC TCT CCA GGG TGT GTC CCT TGC GTT CCG AAG TGT TGG GTG GCC GCA AGG GAC GGC AAA CTC CCT GCA ACG ATC GAT CTG CTT GTC GGA AGC GCC CTC TAC GTG GGG GAC TTG TGC GGG AGT CTC ACG TTC ACG TTC TCC CCC	39 78 117 156 195 234 273 312
35	372577 1		

٥		
	AGG CGC CAC TGG ACA ACG CAA GAC TGT AAC TGT TCT ATC TAT CCC GGC CAC ATA ACG GGT CAC CGC ATG GCA TGG GAT ATG ATG ATG AAC TGG TCC CCC ACA ACA GCC CTG GTA GTA GCT CAG CTG CTC AGG ATC CCG CAA GCC GTC TTG GAC ATG ATC GCT GGT GCC CAC TGG GGA GTC CTA GCG GGC ATA GCG	351 390 429 468 507
5	TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC CTG ATA GTG CTG TTT TCC GGC GTC GAT GCG	546 576
	(2) INFORMATION FOR SEQ ID NO:8:	
10	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(Vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US11	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:	
20	TAC CAA GTA CGC AAC TCC ACG GGG CTT TAC CAT GTC ACC AAT GAT GAT CCC ACA TCG AGT ATT GTG TAC GAG GCG GCC GAT GCC ATC CTG ACC CAC GGG GTGT GTT CCT TGC GTT CCC GGG GGG TGT GTT CCT TGC GTT CCC ACG GTG GCC ACC ACG GTG GCC ACC ACG GTG GCC ACC ACG GTC CAC ACG CAC CTC TCG ACG TCC ACC ACC ACC ACC ACC ACC ACC ACC ACC	39 78 117 156 195 234 273 312 351
25	ATG ATG ATG AAC TGG TCC CCT ACG GCG GCT TTG GTG GTA GCT CAG CTG CTC CGG ATC CCA CAA GCC ATC TTG GAC ATG ATC GCT GGT GCT CAC TGG GGA GTC CTA GCG GGC ATA GCG TAT TTC TCC ATG GTG GGA AAC TGG GCG AAG GTC CTG GTA GCG CTG CTG CTG CTG CTG CTG CTG GCG GTC GAC GCG	429 468 503 546 576
30	(2) INFORMATION FOR SEQ ID NO:9:  (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: D1	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:	

5	AAC GAC CGG CCC GCG GCT TCT CGC TAT ATG TCG GTG	GAC ATG GAG ATA TTC GTT CGG CCC ATG CAG GCG	GTG TGT ATC GAC CTC CGA TGC TTC CAT GGC ATG TTA GGG	TCC ATG AAC GCG CGC TCC GAG CAC AAC CTC GCC	AAC CAC TCC GCT CAC GCC ATC ACG GTG TGG CGG CAC	TCG ACC TCT AGG GTC ATG TCC GTA ACA TCA ATC	AGC CCC CGC AAT TAC CAG CAG CCT CCT CCA GGG	ATT GGG TGC GGC TTG GTG CTG GAG CAC ACA CAA GTC	GTG TGC TGG AAC CTC GGG TTC TGT CGT ACA GCT CTG	TAT GTG GTA GTC GAT ACC AAT ATG GCC GTC GCG	GAG CCC GCG CCC GGG CTC CTC TGC GCT TTA ATG GGC	ACA TGC CTC ACT GCG TGC TCA TGG GTG GAC CTC	GCG GTT ACC ACG GCT GGA CCT ATC GAT GTA ATG GCC	39 78 117 156 195 234 273 312 351 390 429 468 507
10			TCC CTA								GTT.	TTG	ATT.	546 576
15	(2)	(i)			N FOR SEQUE (A) (B) (C) (D) ORIG: (A) (C)	ENCE LENC TYPE STRA TOPO INAL ORGA	CHAI GTH: E: 1 ANDEI OLOGI SOUI	RACTI 576 nucle ONESS Y:	ERIST baseic a S: s linea	se pa acid sing: ar	airs le ens			
		(x:	i)	5	SEQUE	ENCE	DES	CRIP	rion	: SE	Q ID	NO:	LO:	
25	AAT GAC CGG CCC ACA GCT TCT CGC	GAC ATG GAG ACG ATA TTC GTT CGG CCC	GTG TGT ATC GAC CTC CGA TGC TTC CAT GGC ATG	TCC ATG AAC GCG CGC TCC CTC GAG CAC	AAC CAC TCC GCT CAC GCC GTC ACA GTG	TCG ACC TCT AGG GTC ATG TCC GTA ACA	AGC CCC CGC AAT GAT TAC CAG CAG GGT	ATC GGG TGC AGC TTG GTG CTG GAA CAC	GTG TGC TGG AGC CTC GGG TTC TGT CGC	TAT GTG GTA GTC GTT GAT ACC AAC ATG	GAG CCC GCG CCC GGG CTT TTC TGC GCT	ACA TGC CTC ACT GCG TGC TCG TCA TGG	GCG GTT ACC ACG GCT GGA CCT ATC GAT	39 78 117 156 195 234 273 312 351
30	TCG GTG TAC	CAG GCG TAT	TTA GGG TCC CTA	CTC GCC ATG	CGG CAC GTG	ATC TGG GGG	CCA GGG AAC	CAA GTC TGG	GCT CTG GCT	GTC GCG AAG	GTG GGC	GAC CTC	ATG GCC	429 468 507 546 576

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

۰		
	<ul> <li>(A) LENGTH: 576 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
5	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: DK1	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
10	TAT GAA GTG CGC AAC GTG TCC GGG GTG TAC CAC GTC ACA AAC GAC TGC TCC AAC TCA AGC ATC GTG TAT GAG GCA GTG GAC GTG ATC ATG CAT ACC CCA GGG TGC GTG CCC TGC GTT CGG GAC AAC AAC CAC TCC CGT TGC TGG GTA GCG CTC ACC CCC ACG CTC GCG GCC AGG AAC GCC AGC ATC CCC ACT ACG ACA ATA CGA CGC CAT GTC GAT TTG CTC GTT GGG GCG GCT GCT TTC TGC TCC GCT ATG TAC GTG GGG GAC CTC TCC GGT TCC GTT TTC CTC GTC GTC TCC GCT TTC ACC TTT TCA CCT	39 78 117 156 195 234 273 312
15	CGC CGG CAT GAG ACA GCA CAG GAC TGC AAC TGC TCA ATC TAT CCC GGC CAC GTT TCA GGT CAC ACG GAT GCT TGG GAT ATG ATG ATG ATG ATG ACC TGC TCA CAC ACA GCC CTA GTG CTA CCC CAC ATA GCC CTA GTG CTA CCC CAC CAC GCC CTA GTG CAC ATG GCG GGC GCC CAC TGG GGA GTC CTG GCG GGC CTC GCC TAC TAC TAC TCC ATG GCG GGA ACC TGG GCC AAG GTT TTA ATT GCC TTG CTA CTA CTC CTA CCC GCC GCC GCC GCC GCC GCC GCC GCC GC	351 390 429 468 507 546 576
	(2) INFORMATION FOR SEQ ID NO:12:	
20	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
25	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK3	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:	
30	TAT GAA GTG CGC AAC GTG TCC GGG ATA TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC GTC GTG TAT GAG ACA GCA GAC ATG ATG ATG CAT CC CCT GGA TGC GTG GTG CCC TGC GTA CGG GAG AAC ACC TCC CGC TGT TGG GTA GCG CTC ACT CCC ACG CTC GCG GCC AGG AAC GTC ACG GTC CCC ACG CTC ACG GCC TCC GCC TGC TG CTC TGC GGG GCC ACG ACA CACA ATA CGA CGT CAC GCC GCC TCC GCC TCC TCC GCC TCC TCC CCC C	39 78 117 156 195 234 273 312 351
35	TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TGG GAT	390

۰		
	ATG ATG ATG AAC TGG TCC CCT ACA GCA GCC CTA GTG GTG TGG CAA TTA CTC CGG ATC CCG CAA GCT GTC GTG GAC ATG GTG GCG GGG GGC CTA GCG GGA GTC CTA GCG GGC CTA GCC TAC TAT TCC ATG GTG GAA AAC TGG GCT AAG GTT TTG ATT GTG ATG CTA ATG CTA GTG GGC GGC GTT GAT GGG	429 468 507 546 576
5	(2) INFORMATION FOR SEQ ID NO:13:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
10	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK4	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
15	CAT GAA GTG CAC AAC GTA TCC GGG ATC TAC CAT GTC ACG AAC GAC AGC TGC TCC AAC TCA AGT ATT GTG TAT GAG GCA GCG GAC ATG ATC ATG CAT ACC CCC GGG TGC GTG CCC TGC GTC CGG GAG AAC AAC TCC TCC CGT TGC TGG GTA GCG CTC ACT CCC ACG CTC GCG GCA GAC ATC CCC ACT ACG ACA ATA CGC ACG CCC ATG TCC GTT TCC TTC GTT GGG GCG GCT GTC TTC T	39 78 117 156 195 234 273
20	TCT GTC TTC CTC GTC TCC CAG TTG TTC ACC TTC TCG CCT CGC CGG CAT GAG ACG GTA CAG GAC TGC AAT TGC TCA ATC TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GGC TAG GAT ATG ATG ATG ATG ATG ATG ATG ATG	312 351 390 429 468 507 546 576
25	(2) INFORMATION FOR SEQ ID NO:14:	
30	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK5	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
35		

0														
5	AAC GAC CGG CCC GCA GCT TCT CGC TAT ATG TCG GTA TAC	GAC ATG GAA ACG ATA TTC GTT CGA CCC ATG CAG GCG TAT	TGC ATC AAC CTC CGA TGC TTC CAC GGC ATG TTA GGG	TCC ATG AAC GCG CGC TCC CTC GAG CAC AAC CTC GCC ATG	AAC CAC TCC GCC CAC GCT GTC ACG GTA TGG CGG CAC GTG	TTA ACC TCC AGG GTC ATG TCC GTA ACA TCA ATC TGG GGA	AGC CCT CGT AAC GAC TAC CAG CAG GGT CCT CCG GGG AAC	ATC GGG TGT GCC TTG GTG CTG GAC CAC ACA CAA GTC TGG	GTG TGC TGG AGC CTC GGG TTC TGC CGC ACA GCT CTG GCT	TAC GTG GTA GTC GAT ACC AAC ATG GCC GTC GCG AAG	CAT GAG CCC GCG CTT TTC TGC GCT CTA GTG GGC GTT	ACA TGC CTC ACC GCG TGC TCA TGG GTG GAC CTT	ACG GTT GCC ACG GCT GGA CCT ATC GAT GTG ATG GCC	39 788 117 156 195 234 273 312 351 390 429 468 507 546
	(2)	TAIL	FORM	A TT O	I FOI	2 656		NO.	15.					
	(2)	IN	ORM	41101	N FOI	K SEQ	עו ג	NO:	15:					
		(i)	)		EQUI		CHAI			FICS se pa				
					(B)	TYPI		nucle			2115			
15					(C) (D)			ONES:		sing:	le			
					(D)	101	льоо		LINC	41				
		(v:	i)		ORIG (A)	INAL	NOS ISINA		homos	sapi	ens			
					(C)			JAL :			HK	3		
20		(x:	i)		SEQUI	ENCE	DES	CRIP'	rion	: SE	Q ID	NO:	15:	
-0						~=~		~~~		m. a	~ m	am a	3.00	2.0
											CAT			39 78
	GAC	ATG	ATT	ATG	CAT	ACC	CCT	GGA	TGC	ATG	CCC	TGC	GTT	117
											GCG			156
											CCC			195
25											GGG			234 273
23											TTT			312
	CGC	CGA	CAC	GAG	ACG	GTA	CAG	GAC	TGC	AAC	TGC	TCA	ATC	351
											GCT			390
											CTA			429
											GTG GGC			468 507
											GTT			546
30										CCC		110		576

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 576 base pairs

	(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
5	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: IND5	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:	
10	TAT GAA GTG CGC AAC GTG TCC GGG GTG TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGT ATT GTG TAT GAG GCA GCG GAC ATG ATC ACT CCC GGG TGC CCC TGC GTT CCC GAC GAG ACC GCC AGC ACC CTC TCC GCC GCC AGC GTC TCC ACC ACC ACC ACC ACC GCC AGC GTC GTG CCC ACC GCC ACC GTC GTC GTG GTG CCC ACC GCC ACC GTC GAT TTG CTC GCT ACC ACC GCC ACC GTC TCC ACC TCC GCC ACC GTC TCC ACC GCC ACC GTC TCC ACC GCC ACC GTC GTT TCC CCC ACC GCC ACC GTC GTC TCC ACC GTC TCC ACC GTC TCC TCC ACC GCC CGC CGC CAC GTC ATC TCC ACC CAC GCC CTC CCC ACC GTC GCC ACC GTC GTC TCC ACC ACC ACC GCC ACC GCC ACC GTC ATC TCC ACC ACC ACC ACC GCC ACC A	39 78 117 156 195 234 273 312 351 390 429
15	TCG CAG TTG CTC CGG ATC CCA CAA GCT GTC GTG GAT ATG GTG GCG GGG GCC CAC TGG GGA ATC CTG GCG GGC CTT GCC TAC TAT TCC ATG GTA GGG AAC TGG GCT AAG GTT TTG ATT GTG ATG CTA CTC TTT GCC GGC GTT GAC GGG	
	(2) INFORMATION FOR SEQ ID NO:17:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
25	(Vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: IND8	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:	
30	TAT GAG GTG CGC AAC GTG TCC GGG GTG TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGT ATT GTG TAT GAG GCA GCG GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC GTT CGG GAG GGC AAC TTC TCT AGT TGC TGG GTA GCG CTC ACT CCC ACT CTC GCG GCT AGG AAC GCC AGC GTC CCC ACC ACA ATA CGA CGC CAC GTC GAT TTC TC TGT GTC TCT TGT TC TGT GTC TGT TC TGC GGA TCT GTT TC CTT GTC TCC CAC CTG TT ACC TTC TCA CCC CGC CGG CAT GAG ACA GTA CAC GGC GCT TGC AAT TGC TCC TCC TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TCC ATC TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TGC GTA TCC ATC ATC ATC ATC ACC TACA CGC CTA GCT TGC GTA TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TGC GTA TAT CCC GGC CAC GTA TCA GCT ACC CTA GCC CTA GGT GTC GTA TAT CCC GGC CAC GTA TCA GCT ACC CCC CTA GCT GTG GTA	39 78 117 156 195 234 273 312 351 390
35	ATG ATG AAC TGG TCA CCT ACA GCG GCC CTA GTG GTA	429

0		
	TCG CAG TTG CTC CGG ATC CCA CAA GCT GTC GTG GAT ATG GTG GCG GGG GCC CAC TGG GGA ATC CTG GCG GGC CTT GCC TAC TAT TCC ATG GTA GGG AAC TGG GCT AAG GTT TTG ATT GTG ATG CTA CTC TTT GCC GGC GTT GAC GGG	468 507 546 576
5	(2) INFORMATION FOR SEQ ID NO:18:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
10	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: P10	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:	
15	TAT GAA GTG CGC AAC GTG TCC GGG GTG TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGT ATT GTG TAT GAG GCA ACG GAC ATG ATA ATG CAC ACC CCC GGG TGC GTG CCC TGT GTT CGG GAG AAC AAC TCC TCC CGC TGC GG GTA GCG CTC ACT CCC ACA CTC GCG GCT AGG AAT TCC AGC GTC CCA ACT ACG GCA ATA CGA CGC CAT GTC GAT TTG CTC GTT GGG GCG GCT	39 78 117 156 195 234
20	GCT TTC TGC TCC GCT ATG TAC GTG GGG GAT CTC TGC GGA TCT GTT CTC CTC GTC TCC CAG CTG TTC ACC TTC TCA CCT CGC CGG CAT TGG ACA GTA CAG GAC TGC AAT TGT TCA ATC TAT CCT GGC CAC GTA TCA GGT CAC CGC ATG GCT TGG GAT ATG ATG ATG AC TGG TCG CCC ACA GCA GCC CTA GTG GTG TCG CAG CTA CTC CGG ATC CCA CAA GCA GCC CTA GTG GTG GTG GCG GGG GCC CAC TGG GGA GTC CTG GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG ACC TGG GCT ACC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTC TTG ATT GTG ATG CTA CTC TTT GCC GGC GTT GAC GGA	273 312 351 390 429 468 507 576
25	(2) INFORMATION FOR SEQ ID NO:19:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
30	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: S9	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:	
35	TAT GAA GTG CGC AAC GTA TCC GGG GCG TAC CAT GTC ACG	3

0		
5	AAC GAC TGC TCC AAC TCA AGT ATT GTG TAC GAG GCA GCG GAC GTG ATC ATC CCC GGG TGT GTA CCC TGC GTT CAG GAG GGA ACC CTC CAA TGC TGG GTG GGG CTC ACC ACC ACC ACC ACC ACC ACC ACC ACC	78 117 156 195 234 273 312 351 390 429 468 507 546
10		
	(2) INFORMATION FOR SEQ ID NO:20:	
	` '	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs	
	<ul><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: single</li></ul>	
15	(D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE:	
	(A) ORGANISM: homosapiens	
	(C) INDIVIDUAL ISOLATE: S45	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:	
20	TAT GAA GTG CGC AAC GTG TCC GGG GCG TAC CAT GTC ACG	39
	AAC GAC TGC TCC AAC TCA AGC ATT GTG TAT GAG GCA GTG	78 117
	GAC GTG ATC CTG CAC ACC CCT GGG TGC GTG CCC TGC GTT CGG GAG AAC AAC TCC TCC CGT TGC TGG GTG GCG CTC ACT	156
	CCC ACG CTC GCG GCC AGG AAC TCC AGC GTC CCC ACT ACG	195
	ACA ATA CGA CGT CAC GTC GAT TTG CTC GTT GGG GCG GCT GCT TTC TGC TCC GCT ATG TAC GTG GGG GAT CTC TGC GGA	234 273
25	TOT GTT TTC CTT GTT TCC CAG CTG TTC ACC TTC TCG CCT	312
	CGT CGG CAT GAG ACA GTA CAG GAC TGC AAC TGT TCA ATC	351 390
	ATG ATG ATG AAC TGG TCG CCT ACA GCA GCC TTA GTG GTA	429
	TCG CAG TTA CTC CGG ATC CCA CAA GCT GTC GTG GAC ATG	468
	GTG GCG GGG GCC CAC TGG GGA GTC CTG GCG GGC CTT GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT CTG ATT	507 546
20	GTG ATG CTA CTC TTT GCC GGC GTT GAC GGG	576
30		

INFORMATION FOR SEQ ID NO:21: (2)

> SEQUENCE CHARACTERISTICS:
> (A) LENGTH: 576 base pairs
> (B) TYPE: nucleic acid (i)

•										
	<pre>(C) STRANDEDNESS: single (D) TOPOLOGY: linear</pre>									
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA10									
5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:									
10	TAT GAA GTG CGC AAC GTG TCC GGG ATG TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC ATT GTG TAT GAG GCA GCG GAC ATG	78 117 156 195 234 273 312 351 390 429 468								
	GTT ATG CTA CTC TTT GCC GGC GTT GAC GGG	576								
20	(2) INFORMATION FOR SEQ ID NO:22:  (i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear									
25	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SW2									
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:									
30	TAT GAA GTG CGC AAC GTG TCC GGG GTG TAT CAT GTC ACG AAC GAC TGT TCC AAC TCA AGC ATT GTG TAT GAG ACA GCG GAC ARG ATC ATC CAT CAT GCC CCC GGG TGC GTG CCT TGC GTT CGG GAG GCC AAC TCC TCC CGC TGC GTG GTG CCC ACT ACG ACC ATA CGA GCC CAC GTC GAT TTG CTC GTT GGG GCG GCT GCT TTC TGC TCC GTT ATC GAG GCC GCT GCT TTC GTT TTC CTC GTT CCC CAC AGC GCT CACT TTC GTT TTC CTC GTT CCC CAC TTC TCC ACT TTC ACC CCC TCG CAC GAG ACA GTA CAC GCC ACT GCT TTC ACT TTT TCA CCT TAT CCC GCC GAC GAG ACA GTA CAC GAC GCC TTC GTT GCAC TAT CCC GAC TTC TCC CAC CTC TCC CAC CTC TCC TCC CAC ATC ATG ATC ATC TCC CAC CTC TCC CAC CTC TCC CTC GTC GTA TCC CAC TTC TCC CAC CTC ACA GCC CTC GTG GTA TCC CAC TTC TCC CGC ATC CAC GCC CTC GTG GTA TCC CAC TTC TCC CGC ATC CCA CAC GCC CTC GTG GTA TCC CAC TTC TCC CGC ATC CCA CAC GCC CTC GTG GTA TCC CAC TTC CCC CAC ATC CCA CAC GCT GTC GTC GAC ATC	39 78 117 156 195 234 273 312 351 390 429 468								
35										

Ü	GTA GCG GGG GCC CAC TGG GGA GTC CTG GCG GGC CTT GCA TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT	507 546
	GTG ATG CTA CTC TTT GCT GGC GTT GAC GGG	576
	(2) INFORMATION FOR SEQ ID NO:23:	
5	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
10	<ul><li>(vi) ORIGINAL SOURCE:</li><li>(A) ORGANISM: homosapiens</li><li>(C) INDIVIDUAL ISOLATE: T3</li></ul>	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
15	TAC GAA GTG CGC AAC GTG TCC GGG GTG TAC TAT GTC ACG AAC GAC TGT TCC AAC TCA AGC ATT GTG TAT GAC ACA GCG GAC ATG ATC CAC ACC CCT GGG TGC GTG CCC TGC GTT CGG GAG AGC AAT TCC TCC CGC TGC TGG GTA GCG CTT ACT CCC ACG CTC GCG GCC AGG AAC GCC AGC GTC CCC ACT AAC ACA ATA CGA CGT CAC GTC GAC TTG CTG GGT GGC GCT GCT TTC TGT TCC GCT ATG TAC GTG GGG GAT CTC TGC GGA	39 78 117 156 195 234 273
20	TCT GTT TTC CTC GTC TCC CAG CTG TTC ACT TTC TCG CCT CGC CGG CAG GAA ACA GTA CAG GAC TGC AAC TGC TCA TTC TCG GAT TAT CCC GGC CAC GTA ACA GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG ATG ATG TGC CCC ACA ACG GCA CTA GTG GTG CGC CAC CAC GCT GTC GTG GAC ATG GCT GTG GAC ATG GCT GTG GAC ATG GCT GTG GAC ATG GCT GTG GCC GCC CAC ACA GCT GTC GCC GAT CCC ACA GCT GTC GCC GCC TT GCC TAC TAT TCC ATG GTG GGA ACC TGG GCC TT GCC GCC CTT GCC GCC CTA CAC GCT CTC CTC GCC GCC CTT GCC GCC CTC TCC CTC CT	312 351 390 429 468 507 546 576
25	(2) INFORMATION FOR SEQ ID NO:24:	
23	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
30	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: T10	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:	
35	TAT GAA GTG CGC AAC GTG TCC GGG ATG TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC ATT GTG TTT GAG GCA GCG	39 78
	372577_1	

٥								
5	GAC TTG ATC ATG CAC ACC CCC GGG TGC GTG CGG GAG GAG AGC ACC TCC CGC GGG TGC GTG CCC ACG ACG ACG ACG ACG ACC ACG GTC ACG ACG ATC ACG ACG ACG ACG ACG ACG ACG ACG ACG AC	GCG CTC ACT 156 CCC ACT ACG 195 GGG GCG GCT 234 CTC TGC GGA 273 TTC TCG CCT 312 TGC TCA ATC 351 GCT TGG GAC 390 CTA GTG GTG 429 ATG GAC ATG 468 GGC CTT GCC 507 GTT TTA ATT 546						
10								
	(2) INFORMATION FOR SEQ ID NO:25:							
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (Vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: US6							
	(xi) SEQUENCE DESCRIPTION: SE	Q ID NO:25:						
20	TAT GAA GTG CGC AAC GTG TCC GGG ATG TAC AAC GAC TGC TCC AAC TCA AGC ATT GTG TAT GAC ATG ATC ATG ATC ATG CAC ACT CCC GGG TGC GTG GTG GAG AAC AAT TCC TCC CGC TGC TGG GTA CCC ACG CTC GGG GCC AGG AAC GCT AGC GTC ACA ATA CGA CGC CAC GTC GAT TTG CTC GTG GAC TTC TTC TGC TCC GCT ATG TAC GTG GAG GAC	P GAG GCA GCG 78 CCC TGT GTT 117 GCG CTC ACT 156 CCC ACT ACG 195 GGG GCG GCT 234 CCTC TGC GGG 273						
25	TCC GTT TTC CTC ATC TCC CAG CTG TTC ACC CGT CAG CAT GAG ACA GTA CAG GAC TGC AAT TAT CCC GGC CAC GTA TCA GGT CAC CGC ATC ATG ATG ATG AAT TGG TCA CCT ACA GCA GCA TCG CAG TTA CTC CGG ATC CCA CAA GCT GCT GTG GCG GGG GCC CAC TGG GGA GTC CTG GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAC GTG TTG CTA CTC TTT GCC GGC GTT GAC GGC	T TGT TCA ATC 351 G GCT TGG GAT 390 C CTA GTG GTA 429 C ATG GAC ATG 468 G GGC CTT GCC 507 G GTT CTG ATT 546						
30								
	(2) INFORMATION FOR SEQ ID NO:26:							

INFORMATION FOR SEQ ID NO:26: (2)

> (i) SEQUENCE CHARACTERISTICS: LENGTH: 576 base pairs TYPE: nucleic acid STRANDEDNESS: single (A) (B)

(C)

		(D) TOPO:	LOGY:	linear		
	(vi)		NISM:	homosapiens ISOLATE: T2		
5	(xi)	SEQUENCE	DESCRIP	TION: SEQ ID	NO:26:	
10	GCC CAA GTG AC AAC GAC TGT TC GCC GGG GTT C AGG CTG GGA AA CCA AAC GTG GC GGC TTG CGG AC ACG CTC TGC TC GGG GGT ATG CC CGA CGC CAC TC TAC CCC GGT AC ATG ATG ATG ATG ATG ATG GCG GGG GT ATC GCC GGG GGG GGG TAC TTC TCT AC ATC CTC TTG C	C AAT GAG C CAC GTC AT ACA TCC C GTG CGG CT GCC CAC ATC C GCA GCC CT GCC ATC C GCA GCC CC ATC ACT CC TCG CGC GTT CC TCG CT TCG CT TCG CTCG C	AGC ATC CCC GGG CGA TGC GAC ATG TAC GTG CAG ATG CAA GAA GGA CAC CCC ACA GGC GTC GGC GTC	ACC TGG CAG TGG ATA CCG TGG ATA CCG GGC GCT CTT GGTT GTG ATG GGG GAC CTC TGC ATT GTC TGC ATT GTC TGC ATT GTC CCT ATG GCA GCC ACC ATG GCC ACC ATG GCT ATC ATA GCC ACC ATG GGC AAG GTC	CTC GAA	8 7 6 5 4 3 2 1 0 9 8 7 6
20	(2) INFORMAT	(A) LENG (B) TYPE	CHARACT	TERISTICS: 76 base pairs leic acid		
	(vi)	(D) TOPO ORIGINAL (A) ORG	SOURCE	linear		
25	(xi)	SEQUENCE	DESCRI	PTION: SEQ ID	NO:27:	
30	GCA CAA GTG A AAC GAC TGT T GCC GCG GTC C AAA ACG GGA A CCA AAC GTG G GGC TTG CGG A ACG CTC TGC T GCG GTG ATG C CAA CAT CAC T TAC CCT GGC A ATG ATG ATG A GCG TAC GCG A	CT AAT GAC TC CAC GTG CG GTG CGG CG CAC ATT TC GCA GTC GG TTT GTG GG TTT GTG AC ATC ACT AC TGG TGG TG TGG TGG TG TGG TGG TG TGG TG	AGC ATCCC GGC TGC GAC ATCCAG ATCCAG ATCCAG ATCCAG ATCCAG ATCCAG ATCCAG ACCC ACCC	C ACT TGG CAG TGT GTC CCG C TGG ATA CCG C GGC GCC CTC G GTT GTG ATG G GGG GAC CTC G TTC ATC GTC C TGC AAT TGC C CGT ATG GCA G GCC ACC ATC G GTC ATC TTC	CTC CAG 7 TGC GAG 11 GTT TCA 15 ACG CAG 19 TCC GCC 23 TGC GGC 27 TGC GGC 31 TGT ATC 35 TGT GAT 39 ATC CTG 42 AGAC ATC 46	6 5 4 3 2 1 0 9 8
35	GTT AGC GGG G	CA CAC TGG	GGC GI	C AIG IIC GGC	. 110 000 30	. ,

	TAC TTC TCT ATG CAG GGA GCG TGG GCG AAA GTC GTT GTC ATC CTT CTG CTG GCC GCT GGG GTG GAC GCG	546 576
	(2) INFORMATION FOR SEQ ID NO:28:	
5	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
10	<pre>(vi) ORIGINAL SOURCE:</pre>	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
15	GCC GAA GTG AAG AAC ACC AGT ACC AGC TAC ATG GTG ACA AAT GAC TGT TCC AAC GAC AGC ATC ACC TGG CAA CTC CAG GCC GGG GTC CTC CAC GTC CCC GGG TGC GTC CGG GGG AGA GTT GGA AAC GCG TCG CGG TGC TGG ATA CCC GTC CGG CCA AAC GTA GCT GTG CAG CGG CCT GGC GCC CTC ACG CAG GGC TTG CGG ACG CAC ATC GAC ATG GTT GTG ATG TCC GCC ACG CTC TGC TCC GCT CTC TAC GTG GGG GTA TCC GCC GGG GTA ATG CTC GCC GCT CAG ATG TTC ATT ATC TCG CCG CAG CAC CAC TGG TTT GTG CAG GAA TGG AAC TGC TCC ATT	39 78 117 156 195 234 273 312 351
20	TAC CCT GGT ACC ATC ACT GGA CAC CGT ATG GCA TGG GAC ATG ATG ATG ATG ATG TGG TCG CCC ACA ACC ACC ATG ATC TTG GCG TAC GCG TAC GCG GTT CCC GAG GTC ATC ATA GAC ATC ATC AGC GGA GCT CAC TGG GGC GTC ATG TTC GGC CTA GCC TAC TTC TCT ATG CAG GGA GCG TGG GCG AAG GTC GTT GTC CTG TTG CTC ACC GCT GGC GTG GAC GCG	390 429 468 507 546 576
	(2) INFORMATION FOR SEQ ID NO:29:	
25	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
30	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US10	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:	
35	GTC CAA GTG AAA AAC ACC AGT ACC AGC TAT ATG GTG ACC AAT GAC TGC TGC AAC GAC AGC ATC ACT TGG CAA CTT GAG GCT GCG GTC CTC CAC GTT CCC GGG TGT GTC CCG TGC GAG	39 78 117

٥															
5	AAA CCA GGC ACG GGG CGC TAC ATG GCG ATT TAC	AAT TTG CTC ATG CAC CCC ATG TAC AGC TTC	GTG CGG TGC ATG CAC GGT ATG GTG GGG TCT	GCC ACT TCC TCG ACC AAC ATG GCG ATG	GTG CAC GCT GCA TTT ATC TGG CGC CAT CAG	CAG ATC CTT GCC GTG ACC TCG GTT TGG GGA	CGG GAC TAC CAA CAG GGG CCC CCC GGC GCC	CCT ATG GTG ATG GAA CAC ACG GAG GTC TGG	GGC GTC GGG TTC TGC CGT GCC GTC TTG GCG	GCC GTG GAC ATT AAC ATG ACT ATC TTC	CTC ATG TTC GTC TGC GCA TTG ATA GGC	ACG TCC TCG TCC TCC TGG ATC GAC TTA	CAG GCC GGT CCG ATC GAC CTG ATC GCC		156 195 234 273 312 351 390 429 468 507 546 576
10	(2)	INE	ORMA	OITA	I FOI	R SE	Q ID	NO:3	30:						
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: DK8														
		(x:	i)		SEQU.	ENCE	DES	CRIP'	TION	: SE	Q ID	NO:	30:		
20	AAT GAC AAT CCT AAC	GAT GCA GAC AAT CTG	TGC GTT AAT GTG CGA	TCA CTC GGC GCT ACA	AAC CAC ACC GTG CAC	AAC CTT CTG AAA GTC	AGC CCC CGC CAC GAC	ATC GGA TGC CGC GTG	TGC TGG GGC ATC	TGG GTC ATA GCA GTA	TAC CAA CCA CAA CTT ATG GTA	TGT GTG ACT GCA	ACC GAG ACA CAT GCT		39 78 117 156 195 234 273
25	GCC GAA TAC ATG GCC GTC TAT	GTG CGC CAA ATG TAT TTC TTC	ATG CAC GGT CTA GCC GGC TCC	ATC AAC CAT AAC GCT GGC ATG	GTG TTT ATC TGG CGT CAT CAG	TCG ACC ACC TCA GTT TGG GGA	CAG CAG GGC CCA CCT GGC	GCT GAG CAC ACT GAG GTG	CTC TGC CGC CTT CTA GTG GCC	ATA AAC ATG ACC GCC TTT AAA	ATA TGT GCA ATG CTC GGC GTC	TCG TCC TGG ATC CAG TTG	CCT ATC GAC		312 351 390 429 468 507 546

#### (2) INFORMATION FOR SEQ ID NO:31: 30

(i) SEQUENCE CHARACTERISTICS:

- LENGTH: 576 base pairs TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear (A)
- (B)
- (C)
- (D)

	<ul><li>(vi) ORIGINAL SOURCE:</li><li>(A) ORGANISM: homosapiens</li><li>(C) INDIVIDUAL ISOLATE: DK11</li></ul>							
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:31:						
5	AAT GAT TGC TCA AAC GCA GTT CTC AAT GAC AAT GGC CCT AAT GTG GCC AAC CTG CGA GCC GCC GTG ATG ATG	G AAC ACC AGT TCT AGT TAC TAC GCC ACC A AAC AAC AGC ATC ACC TGG CAA CTC ACC CAC CTT CCC GGA TGC GTC CCA TGT GAG 11' C ACC CTG CAC TGC TGG ATA CAA GTG ACA T GTG AAA CAC CGC GGC GCA CTC ACT CAC A CAT ATA GAT ATG ATT GTA ATG GCA GCT G GCC TTG TAT GTG GGA GAC GTG TGC GGG C TTT ACC CAA GAG TGC AAC GTT TCC ATC 35:						
10	TAC CAA GGT CAC ATG ATG CTT AAC GCC TAT GCC GCC GTC TTC GGT GGT TAT TTC TCC ATC	C TTT ACC CAA GAG TGC AAC TGT TCC ATC C ATC ACC GGC CAC CGC ATG GCA TGG GAC C TGG TCA CCA ACT CTC ACC ATG ATC CTC C GGT GTT CCT GAG CTA GTC CTT GAA GTC T CAT TGG GGT GTG GTG TTT GGC TTG GCC G CAG GGA GCG TGG GCC AAG GTC ATT GCC T GTA GCA GGA GTG GAT GCA T GTA GCA GGA GGA GTG GAT GCA  57:						
15	(2) INFORMATION	ON FOR SEQ ID NO:32:						
20	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE:							
		(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SW3						
25	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:32:  GG AAC ATC AGT TCT AGC TAC TAT GCC ACC  3						
	AAT GAT TGC TC AAC GCA GTC CT AAT GAT AAT GG CCT AAT GTG GC AAC CTG CGA GC	AAC AGC AGC ATC ACC TGG CAA CTC ACC TC CAC CTT CCC GGA TGC GTC CGG TGT GAG 11 CTC GTG CAC TGC TGG ATA CAA GTG ACA TC GTG CAC TGC TGG ATA CAA GTG ACA TC GTG AAA CAC CGC GGC GCC CTC ACT CAC TACAC GTC GAT ATG ATC GTA ATG GCA GCT CAC GTC GTT GTG GGA GAC ATG TGC GGG TC GTG TCG CAG GCT TTC ATA ATA TCG CCA 31						
30	GAA CGC CAC AA TAC CAA GGT CG ATG ATG CTA AA GCC TAT GCC GG GTC TTC GGC GG TAT TTC TCC AT	AC TTT ACC CAA GAG TGC AAC TGT TCC ATC 35 37 ATC ACC GGC CAC CGC ATG GGC TGG GAC 36 37 ATC ACC GGC CAC GGC ATG GGC TGG GAC 37 ACC TGG TCA CCA ACT CTT ACC ATG ATC CTT 37 CGT GGT CCT GAG CTA GTC CTT GAA GTT 38 GC CAT TGG GGC GTG GTG TTT GGC TTG GCC 39 GCA GGA GGC GG GCA GGC ATG GCC 40 GCA GGA GGC GGA GGC 50 GGA GGC GGA GGC 51 GGC GCA GGA GGC GCA 51 GGC GCA GGA GGC GCA 52 GGC GGA GGA GGC GCA 53 GGC GGA GGC GGA GGC 54 GGC GGA GGC GCA 55 GGC GGA GGC GGC GCA 56 GGC GGA GGC GGC GCA 57 GGC GCA GGA GGC GCA 57 GGC GGC GGC GGC GGC GGC GGC GGC 57 GGC GGC GGC GGC GGC GGC GGC GGC GGC GG						
35								

	(2) INFORMATION FOR SEQ ID NO:33:	
5	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	<ul><li>(vi) ORIGINAL SOURCE:</li><li>(A) ORGANISM: homosapiens</li><li>(C) INDIVIDUAL ISOLATE: T8</li></ul>	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
15	GTG GAA GTT AGA AAC ACC AGT TTT AGC TAC TAC GCC ACC AAT GAT TGC TCG AAC AAC AGC ATC ACC TGG CG CTC ACC AAC GCA ATC GCC GGA TGC GCC TGC CAC ACC CTC CCC GGA TGC GCC CCA TGT GAC AAT GAC AAT GGC ACC TTG CGC TGC TGG ATA CAA GTA ACA CCT AAT GTG GCT GTC AAA CAC CTG GCG GCA CTC ACT CAC AAC CTG ACG CAT GTC GAC GTG ATC GTA ATG GCA GCT ACC GCC TGC GCC GCC GCC GCC GCC GCC GCC G	39 78 117 156 195 234 273
20	GCC GTG ATG ATA GCS TCG CAS GCT TTC ATA ATA TCG CCA GAA CGC CAC AAC TTC ACC CAS GAS TGC AAC TGT TCC ATC TAC CAA GGT CAT ATC ACC GGC CAC CGC ATG GCA TGG GAC ATG ATG CTG AAC TGG TCA CCA ACT CTC ACC ATG ATC CTC GCC TAC GCT GCT CGT GTG CCT GAA CTA GTC CTT GAA GTT GTC TTC GGC GGC CAT TGG GGC GTG GTG TTT GGC TTG TAT TTC TCC ATG CAA GGA GCC TGG GCC AAA GTC ATC GCC TAT TTC TCC ATG CAA GGA GCC TGG GCC AAA GTC ATC GCC	312 351 390 429 468 507 546
20	ATC CTC CTC GTC GCA GGA GTG GAC GCA  (2) INFORMATION FOR SEQ ID NO:34:	576
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: S83	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
	GTG GAG GTC AAG GAC ACC GGC GAC TCC TAC ATG CCG ACC AAC GAT TGC TCC AAC TCT AGT ATC GTT TGG CAG CTT GAA GGA GCA GTG CTT ACT ACT CCT GGA TGC CCT TGT GAG CGT ACC GCC AAC GTC TCT CGA TGT TGG GTG CCC GTT GCC CCC AAT CTC CCA AGC CATA AGT CAA CCT GGC GCT CTC ACT AAG	39 78 117 156 195
35		

0		
5	ACG GTC TGT TCT GCC GCG CTG ATG CTG GCC CAA CAC CAT ACG TTT TAC CCG GGC CGC ATT ATG ATG ATG AAC TGG GCG TAC TTG GTG CGC GTT ACA GGA GGT CAT	ATC GAT ATC ATC GTG ATG TCT GCT CTT TATT GTG GGG GAC GTG TGT GGC GCT CAG GTC GTC GTC GTG TCG CCA GTC CAG GAA TGC AAC TGT TCC ATA ACG GGA CAC CGC ATG GCT TGG GAT TCG CCC ACT ACC ACC ATG CTC TGG GAT ATC CCG GAA GTC ATC TTG GAT ATT ATG GGT GTA ATG TTT GG CTC CTG GGA TCG TGG GAG GTC ATC GTT GG GT GTA ATG TTT GG CTC GCT GGA TCG TGG GCG AAG GTC ATC GTT 50' GGA TCG TGG GCG AAG GTC ATC GTT 50' GCT GGG GTG GAG GCG 570'
	(2) INFORMATION FOR	SEQ ID NO:35:
10	(A) (B) (C)	NCE CHARACTERISTICS: LENGTH: 576 base pairs TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear
15	(A) (C)	NAL SOURCE: ORGANISM: homosapiens INDIVIDUAL ISOLATE: DK12 ENCE DESCRIPTION: SEQ ID NO:35:
	(xi) SEQUE	INCE DESCRIPTION: SEQ ID NO:35:
20	AAC GAC TGT TCC AAT GAC GTC ATT CTG CAC CAG GAC GGC AAT ACA CCT ACA GTG GCA GTC TCG ATA CGC AGT CAT ACG ATG TGC TCT GCG GCC GTC TTC CTT GTG	GTG TCC GGC CTC TAC GTC CTT ACC AGC AGT ATC GTG TAT GAG GCC GAT ACA CCT GGC TGT GTA CCT TGT GTT TCT ACG TGC TGG ACC TCA GTG ACG AGG TAC GTC GGA GCA ACC ACC GCT GTG GAC CTG CTA GTG GGC GCC CTC TAC GTG GGT GAT GTG TGT GGC CTC TAC GTG GGT GAT GTG TGT GGG GGA CAA GCC TCT ACG TCC ACG CTC CAG ACC TCT ACG TCC TCG CTG GGA CAA GCC TTC ACG TCC TCG CTG TC CAG ACC TGT AAC TCC TCG CTG
25	TAC CCA GGC CAT CTT ATG ATG ATG AAT TGG GCG CAC GTC CTG CGT ATA GCT GGG GCC CAT	TCA GGA CAT CGA ATG GCT TGG GAT 39 TCC CCC GCT GTG GGT ATG GTG GTA 42 CTG CCC CAG ACC TTG TTC GAC ATA 46 TGG GGC AAC TGG GCC GAG GCC TAG GCC GGC AAC TGG GCC AAG GTC GCT ATC 54
	(2) INFORMATION FOR	R SEQ ID NO:36:
30	(i) SEQUI (A) (B) (C) (D)	ENCE CHARACTERISTICS: LENGTH: 576 base pairs TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear

ORIGINAL SOURCE:
(A) ORGANISM: homosapiens

35

(vi)

# (C) INDIVIDUAL ISOLATE: HK10

(xi)	CECTIENCE	DESCRIPTION:	SEO	TD	NO:36:
(X1)	SECOFICE	DESCRIPTION.	SEQ	TD	140.50.

CTA GAG TGG CGG AAT GTG TCT GGC CTC TAT GTC CTT ACC ACC ACC ACC ACC ACC ACC ACC A															
5 GAC GTC ATT CTG CAC ACA CCT GGC TGT GTA CCT TGT GTT  CAG GAC GGC AAT ACA TCC ACG TGC TGT GTA CCT TGT GTT  CCT ACA GTG GCA GTC AGG TAC GTC TGG ACC TCG GTG ACA  CCT ACA GTG GCA GTC AGG TAC GTC GGC GCG GCC L95  TCG ATA CGC AGT CAT GTG GAC CTG TTA GTG GGC GCC 234  ACG ATG TGC TCT GCG CTC TAC GTG GGC ATAT GTG GGC GCC  GCC GTC TTC CTC GTG GGA CAA GCC TTC ACG TTG AGA CCG 312  CGT CGC CAT CAA ACG GTC CAG ACC TGT ACA TGC TGG GTG 351  TAC CCA GGC CAC CTT TCA GGA CAT CGA ATG GCT TGG GAT 390  ATG ATG ATG AAT TGG TCC CCC GCC GTG GGT ATG GTG GTG 429  GCG CAC GTC CTG CGG TTG CCC CCC GCC GTG GGT ATG GTG GTG 429  GCG CAC GTC CTG CGG GGC ATC TTG GCA CTT TGC GCC GCC GTG GGT ATG GTG GTG 429  ATA GCC GGG GCC CAT TGG GGC ATC TTG GCA GGC CTA GCC 507  TAT TAC TCC ATG CAG GGC AAC TTG GCA AGG GTC GTA TCC  TAT TAC TCC ATG CAG GGC AAC TTG GCA AGG GTC GCT ATC		CTA	GAG	TGG	CGG	AAT	GTG	TCT	GGC	CTC	TAT	GTC	CTT	ACC	
5 GAC GTC ATT CTG CAC ACA CCT GGC TGT GTA CCT TGT GTT  117 CAG GAC GGC AAT ACA TCC ACG TGC TGG ACC TCG GTG ACA  CCT ACA GTG GCA GTC AGG TAC GTC GGA GCA ACC ACC GCC  TCG ATA CGC AGT CAT GTG GAC CTT TTA GTG GGC GCC  ACG ATT CGC CTC TAC GTG GGC GAT ATG TGT GGC  GCC GTC TTC CTC GTG GGA CAA GCC TTC ACG TTC AGA CCG  CGT CGC CAT CAA ACG GTC CAG ACC TGT AAC TGC CTG  TAC CAC AGC CAC CTT TCA GGA CAT CGA ATG GTC GGG  ATG ATG ATG ATT GTG CC CCC GCC GTG GGT ATG GTG GTG  ATG ATG ATG ATT GTG CCC CCC GCC GTG GGT ATG GTG GTG  ATG ACG GCC CAT TGG GGC ATC TTG CCC CCC GCC GTG GGT ATG GTG GTG  ATG ACG GCC CAT CAA ACC GTC CCC CCC ACC GTG GCT ATG GTG GTG  ATG ATG ATG ATT ACT GCC CAC ACC TTT GCC AGG CTTA ACC  TAT TAC TCC ATG CAG GGC AAC TTG GCA AGG GCC CTA GCC  TAT TAC TCC ATG CAG GGC AAC TGG GCC AAG GTC GCT ATC		AAC	GAC	TGT	CCC	AAT	AGC	AGT	ATT	GTG	TAT	GAG	GCC	GAT	78
CAG ACA GGC AT ACA FACE RC 160 FOR 160 ACC ACC GCC CT ACA GTG GCA GCA ACC ACC GCC ACC ACC ACC ACC AC	_	GAC	GTC	ATT	CTG	CAC	ACA	CCT	GGC	TGT	GTA	CCT	TGT	GTT	117
CCT ACA GTG GCA GTC AGG TAC GTC GGA GCA ACC ACC GCC   195	5	CAG	GAC	GGC	AAT	ACA	TCC	ACG	TGC	TGG	ACC	TCG	GTG	ACA	156
TCG ATA CGC AGT CAT GTG GAC CTG TTA GTG GGC GCG GCC															195
ACG ATG TGC TCT GCG CTC TAC GTG GGC GAT ATG TGT GGG 273 GCC GTC TTC CTC GTG GGA CAA GCC TTC ACG TCT AGA CCG 312 CGT CGC CAT CAA ACG GTC CAG ACC TGT AAC TGC TCG CTG 351 TAC CCA GGC CAC CTT TCA GGA CAT CGA ATG GCT TGG GAT 390 ATA GAT ATG AAT TGG TCC CCC GCC GTG GGT ATG GTG GTG 429 GCG CAC GTC CTG CGG TTG CCC CAC ACC TTG TCC GAC ATA 468 ATA GCC GGG GCC CAT TGG GGC ATC TGG ACC GTG GCC TAG GCC TAG CCC TAT TAT TAC TCC ATG CAG GGC AAC TGG GCC AAG GCC CTA GCC 507															234
CGT CGC CAT CAA ACG GTC CAG ACC TGT AAC TGC TCG CTG															273
CGT CGC CAT CAA ACG GTC CAG ACC TGT AAC TGC TCG CTG   351		GCC	GTC	TTC	CTC	GTG	GGA	CAA	GCC	TTC	ACG	TTC	AGA	CCG	312
10 ATG ATG ATG AAT TGG TCC CCC GCC GTG GGT ATG GTG GTG GCG CAC GTC CTG CGG TTG CCC ATG CTG TTG TTC GAC ATA ATA GCC GGG GCC CAT TGG GGC ATC TTG GCA GGC CTA GCC TAT TAC TCC ATG CAG GGC AAC TGG GCC AAG GTC GCT ATC 546															
GCG CAC GTC CTG CGG TTG CCC CAG ACC TTG TTC GAC ATA ATA GCC GGG GCC CAT TGG GGC ATC TTG GCA GGC CTA GCC TAT TAC TCC ATG CAG GGC AAC TGG GCC AAG GTC GCT ATC  546		TAC	CCA	GGC	CAC	CTT	TCA	GGA	CAT	CGA	ATG	GCT	TGG	GAT	
GCG CAC GTC CTG CGG TTG CCC CAG ACC TTG TTC GAC ATA 468 ATA GCC GGG GCC CAT TGG GGC ATC TTG GCA GGC CTA GCC 507 TAT TAC TCC ATG CAG GGC AAC TGG GCC AAG GTC GCT ATC 546	10	ATG	ATG	ATG	AAT	TGG	TCC	CCC	GCC	GTG	GGT	ATG	GTG	GTG	429
TAT TAC TCC ATG CAG GGC AAC TGG GCC AAG GTC GCT ATC 546	10	GCG	CAC	GTC	CTG	CGG	TTG	CCC	CAG	ACC	TTG	TTC	GAC	ATA	
TAT TAC TCC ATG CAG GGC AAC TGG GCC AAG GTC GCT ATC 546		ATA	GCC	GGG	GCC	CAT	TGG	GGC	ATC	TTG	GCA	GGC	CTA	GCC	507
ATC ATG GTT ATG TTT TCA GGG GTC GAT GCC 576															546
		ATC	ATG	GTT	ATG	TTT	TCA	GGG	GTC	GAT	GCC				576

### (2) INFORMATION FOR SEQ ID NO:37: 15

- SEQUENCE CHARACTERISTICS: (i)
  - (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- ORIGINAL SOURCE: (vi) 20
  - (A) ORGANISM: homosapiens(C) INDIVIDUAL ISOLATE: S2

#### SEQUENCE DESCRIPTION: SEQ ID NO:37: (xi)

25	CTA	GAG	TGG	CGG	AAT	ACG	TCT	GGC	CTC	TAT	GTC	CTC	ACC	
	AAC	GAC	TGT	TCC	AAT	AGC	AGT	ATT	GTG	TAT	GAG	GCC	GAT	
	GAC	GTT	ATT	CTG	CAC	ACA	CCT	GGC	TGT	GTA	CCT	TGT	GTT	
							ACG							
	CCT	ACA	GTG	GCA	GTC	AGG	TAT	GTC	GGA	GCA	ACC	ACC	GCT	
							GAC							
30							TAC							
	GCC	GTC	TTT	CTC	GTG	GGA	CAA	GCC	TTC	ACG	TTC	AGA	CCT	
	CGT	CGC	CAT	CAA	ACG	GTC	CAG	ACC	TGT	AAC	TGC	TCG	CTG	
	TAC	CCA	GGC	CAT	CTT	TCA	GGA	CAT	CGC	ATG	GCT	TGG	GAT	
							CCC							
	GCG	CAC	GTT	CTG	CGT	TTG	CCC	CAG	ACC	GTG	TTC	GAC	ATA	
	ATA	GCC	GGG	GCC	CAT	TGG	GGC	ATC	TTG	GCG	GGC	CTA	GCC	
	TAT	TAC	TCC	ATG	CAA	GGC	AAC	TGG	GCC	AAG	GTC	GCT	ATC	
	ATC	ATG	GTT	ATG	TTT	TCA	GGG	GTC	GAC	GCC				

0		
	(2) INFORMATION FOR SEQ ID NO:38	3:
5	(i) SEQUENCE CHARACTER (A) LENGTH: 576 (B) TYPE: nuclei (C) STRANDEDNESS: (D) TOPOLOGY: 1:	base pairs ic acid : single
,	(vi) ORIGINAL SOURCE:  (A) ORGANISM: h  (C) INDIVIDUAL IS	SOLATE: S52
	(xi) SEQUENCE DESCRIPT	ION: SEQ ID NO:38:
10	CTA GAG TGG CGG AAT ACG TCT GGC (AAC GAC TGT TCC AAT AGC AGT ATT GG AC GGC AC GC AC GGC AC	GTG TAT GAG GCC GAT 78 TGT GTA CCT TGT GTT 117 TGG ACC CCA GTG ACA 156 GGA GCA ACC ACC GCT 195 TTA GTG GGC GCC GCC 234 GGT GAT ATG TGT GGG 273 TTC ACG TTC AGA CCT 312
15	CGT CGC CAT CAA ACG GTC CAG ACC TAC CCA GGC CAT GTT TCA GGA CAT ATG ATG ATG ATG TCC CCC GGT GCG CAC ATC CGG GCC CAT GG GGC ATC ATG TCC ATG ATG TCC ATG GCC ATG	TGT AAC TGC TCG CTG 351 CGA ATG GCT TGG GAT 390 GTG GGT ATG GTG GTG 429 ACC TTG TTT GAC ATA 468 TTG GCG GGC CTA GCC 507 GCC AAG GTC GCT ATT 546
20		
	(2) INFORMATION FOR SEQ ID NO:3	
25	(i) SEQUENCE CHARACTE (A) LENGTH: 576 (B) TYPE: nucle (C) STRANDEDNESS (D) TOPOLOGY: 1	base pairs ic acid : single
	(-/ -	SOLATE: S54
	(xi) SEQUENCE DESCRIPT	CION: SEQ ID NO:39:
30	CTA GAG TGG CGG AAT ACG TCT GGC AAC GAC TGT TCC AAT AGC AGT ATT GAC GTC ATT GTG CAC ACA CCC GGC CAG GAC GGC AAT ACA TCC ACG TGC CCT ACG GGC AAT ACA TCC ACG TGC TCG ATA CGC AGT CAT GTG GAC CTA ACG CTG TGC TCT GCG CTC TAT GTG	GTG TAT GAG GCC GAT TGT GTA CCT TGT GTT TGG ACC CCA GTG ACC GGA GCA ACC ACC GCT TTA GTG GGC GCC  234

372577\_1

GCC GTC TTT CTC GTG GGA CAA GCC TTC ACG TTC AGA CCT 312 CGT CGC CAT CAA ACG GTC CAG ACC TGT AAC TGC TCG CTG 351 TAC CCA GGC CAT CTT TCA GGA CAT CGA ATG GCT TGG GAT 390 ATG ATG ATG AAT TGG TCC CCC GCT GTG GGT ATG GTG GTG 429 GCG CAC ATC CTG CGA TTG CCC CAG ACC TTG TTT GAC ATA 468 CTG GCC GGG GCC CAT TGG GGC ATC TTG GCG GGC CTA GCC 507 TAT TAT TCT ATG CAG GGC AAC TGG GCC AAG GTC GCT ATC 546 5 ATC ATG ATT ATG TTT TCA GGG GTC GAT GCC 576 INFORMATION FOR SEQ ID NO:40: (2) SEQUENCE CHARACTERISTICS: (i) LENGTH: 576 base pairs (A) TYPE: nucleic acid (B) 10 STRANDEDNESS: single (C) TOPOLOGY: linear (D) (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens INDIVIDUAL ISOLATE: (C) 15 SEQUENCE DESCRIPTION: SEQ ID NO:40: (xi) GAG CAC TAC CGG AAT GCT TCG GGC ATC TAT CAC ATC ACC 39 AAT GAT TGT CCG AAT TCC AGT ATA GTC TAT GAA GCT GAC 78 CAT CAC ATC CTA CAC TTG CCG GGG TGC GTA CCC TGT GTG 117 ATG ACT GGG AAC ACA TCG CGT TGC TGG ACG CCG GTG ACG 156 CCT ACA GTG GCT GTC GCA CAC CCG GGC GCT CCG CTT GAG 195 TCG TTC CGG CGA CAT GTG GAC TTA ATG GTA GGC GCG GCC 234 20 ACT TTG TGT TCT GCC CTC TAT GTT GGG GAC CTC TGC GGA 273 GGT GCC TTC CTG ATG GGG CAG ATG ATC ACT TTT CGG CCG 312 CGT CGC CAC TGG ACC ACG CAG GAG TGC AAT TGT TCC ATC 351 TAC ACT GGC CAT ATC ACC GGC CAC AGG ATG GCG TGG GAC 390 ATG ATG ATG AAC TGG AGC CCT ACC ACC ACT CTG CTC CTC 429 GCC CAG ATC ATG AGG GTC CCC ACA GCC TTT CTC GAC ATG 468 GTT GCC GGA GGC CAC TGG GGC GTC CTC GCG GGC TTG GCG 507 TAC TTC AGC ATG CAA GGC AAT TGG GCC AAG GTA GTC CTG 546 25 GTC CTT TTC CTC TTT GCT GGG GTA GAC GCC 576 INFORMATION FOR SEQ ID NO:41: (2)

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 576 base pairs

(A) LENGTH: 576 base pairs (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: Z1

35

### SEQUENCE DESCRIPTION: SEQ ID NO:41: (xi) GTG CAC TAC CGG AAT GCT TCG GGC GTC TAT CAT GTC ACC 39 AAT GAT TGC CCT AAC ACC AGC ATA GTG TAC GAG ACG GAG 78 CAC CAC ATC ATG CAC TTG CCA GGG TGT GTC CCC TGT GTG 117 CGG ACG GAG AAT ACT TCT CGC TGC TGG GTG CCC TTG ACC 156 CCC ACT GTG GCC GCG CCC TAT CCC AAC GCA CCG TTA GAG 195 5 TCC ATG CGC AGG CAT GTA GAC CTG ATG GTG GGT GCG GCT 234 ACT ATG TGT TCC GCC TTC TAC ATT GGA GAT CTG TGT GGA 273 GGC GTC TTC CTA GTG GGC CAG CTG TTC GAC TTC CGA CCG 312 CGC CGG CAC TGG ACC ACC CAG GAT TGC AAC TGC TCC ATC TAT CCT GGT CAC GTC TCG GGC CAC AGG ATG GCC TGG GAC 390 ATG ATG ATG AAC TGG AGC CCT ACC AGC GCG CTG ATT ATG 429 GCT CAG ATC TTA CGG ATC CCC TCT ATC CTA GGT GAC TTG 468 CTC ACC GGG GGT CAC TGG GGA GTT CTT GCT GGT CTA GCT 507 10 TTC TTC AGC ATG CAG AGT AAC TGG GCG AAG GTC ATC CTG 546 GTC CTA TTC CTC TTT GCC GGG GTC GAG GGA 576 INFORMATION FOR SEQ ID NO:42: (2) SEQUENCE CHARACTERISTICS: (i) 15 LENGTH: 576 base pairs (A) (B) TYPE: nucleic acid STRANDEDNESS: single (C) TOPOLOGY: linear (D) (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens INDIVIDUAL ISOLATE: (C) 20 SEQUENCE DESCRIPTION: SEQ ID NO:42: (xi) GTT AAC TAT CGC AAT GCC TCG GGC GTC TAT CAC GTC ACC AAC GAC TGC CCG AAC TCG AGC ATA GTG TAT GAG GCC GAA CAC CAG ATC TTA CAC CTC CCA GGG TGC TTG CCC TGT GTG 117 AGG GTT GGG AAT CAG TCA CGC TGC TGG GTG GCC CTT ACT 156 CCC ACC GTG GCG GTG TCT TAT ATC GGT GCT CCG CTT GAC 195 25 TCC CTC CGG AGA CAT GTG GAC CTG ATG GTG GGC GCC GCT 234 ACT GTA TGC TCT GCC CTC TAC GTT GGA GAT CTG TGC GGT 273 GGT GCA TTC TTG GTT GGC CAG ATG TTC TCC TTC CAG CCG 312 CGA CGC CAC TGG ACT ACG CAG GAC TGC AAT TGT TCT ATC TAC GCA GGG CAT ATC ACG GGC CAC AGG ATG GCA TGG GAC 390 ATG ATG ATG AAC TGG AGT CCC ACA ACC ACC CTG CTT CTC 429

GCC CAG GTC ATG AGG ATC CCT AGC ACT CTG GTA GAT CTA

CTC GCT GGA GGG CAC TGG GGC GTC CTT GTT GGG TTG GCG

TAC TTC AGT ATG CAA GCT AAT TGG GCC AAA GTC ATC CTG

468

507

546 576

(2) INFORMATION FOR SEQ ID NO:43:

GTC CTT TTC CTC TTC GCT GGA GTT GAT GCC

35

0		
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
5	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: Z7	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:	
10	GTC AAC TAT CAC AAT GCC TCG GGC GTC TAT CAC ATC ACC AAC GAC TGC CCG AAC TCG AGC ATA ATG TAT GAG GCC GAA CAC CAC CAC ATC CTA CAC CTC CCA GGG TGC GTA CCC TGT GTG AGG GAG GGG AAC CAG TCA CGC TGC TGG GTG GCC CTT ACT CCC ACC GTG GCG GCG CCT TAT ATC GGT GCA GCC TCT GAA ACA GTG TGC CCC CCC CCC TCC TCC TCC ATC GCG GCT CCC ACC GCT TCC ATC GCC TCC CCC ACC GCT CCC ACC ACC GCT CCC ACC ACC GCT CCC ACC ACC ACC GCT CCC ACC ACC ACC ACC ACC ACC ACC ACC A	39 78 117 156 195 234 273 312
15	CGA CGC CAC TGG ACT ACG CAG GAC TGC AAT TGT TCC ATC TAT GCG GGG CAC GTT ACA GGC CAC AGA ATG GCA TGG GAC ATG ATG ATG AAC TGG ACT CCC ACA ACC ACC TTG GTC CTC GCC CAG GTT ATG AGG ATC CCT AGC ACT CTG GAG CTA CTC ACT GGA GGC CAG GCA TAG GAT ATT GG GGT ATC CTT ATC GGG GTG GCA TGG GCA TGG CTT ATC GGG GTG GCA TGC CTT ATC GGG GTG CAT CTT ATC GGG GTG GCA TGC CTT ATC GGG GTG CAT CTT ATC GGG GTG GCA TGC CTT TTC CTC TAC GCT GGA GTT GAT GCC	351 390 429 468 507 546 576
20	(2) INFORMATION FOR SEQ ID NO:44:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
25	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: DK13	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:	
30	TAC AAC TAT CGC AAC AGC TCG GGT GTC TAC CAT GTC ACC AAC GAT TGC CCG AAC TCG AGC ATA GTC TAT GAA ACC GAT TAC CAC ATC TTA CAC CTC CCG GGA TGC GTT CCT TCG GTG AGG GAA GGG ATA GTC TCT CTC ACC CCC ACC GTG GCT AAC ATC TCT ACA CTC TGA TGC TCT TTG AGA CGT CAC GTG GAA CAT CTG ATT GGT GGC GGC ACT TCT TTG AGA CGT CAC GTG GAT ATC TCT TTG AGA CGT CAC GTG GAT ATC ATC GGT GGC GGC GCC ACT CTC TCC GCC CTC TAC ATC GGT GAG GAC GTT GTG GGG GGT GTT TTT GGTC GGT CAC CTC TAC ACC GTG TTC ACC TTC CAA CCT	39 78 117 156 195 234 273 312
35	CGC CGC CAC TGG ACC ACC CAA GAC TGC AAT TGT TCC ATC	351

	- 107 -	
۰		
5	TAC ACA GGA CAT ATC ACA GGA CAC AGA ATG GCT TGG GAC ATG ATG ATG ATT TGG AGC CCC ACT GGG ACG CTG GTC CTC GCC CAA CTT ATG AGG ATC CCA GGC GCC ATG GTC CTG CTT GCA GGC GCC CATG GTC CTG GCC ATG GCC ATG GCC ATG CTG CTG TCC TTC AGC ATG CAA GCT AAT TGG GCC AAG GTT ATC CTG GTC CTG TTT CTC TTT GCT GGA GTC GAC GCT	390 429 468 507 546 576
J	(2) INFORMATION FOR SEQ ID NO:45:	
10	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: SA1	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:	
15	GTT CCC TAC CGG AAT GCC TCT GGG GTT TAC CAT GTC ACC AAT GAC TGC CCA AAC TCC TCC ATA GTC TAC GAG GCT GAT AGC CTG ATC TTG CAC GCA CCT GGC TGC GTG CCC TGT GTC AGG CAA GAT AAT GTC AGT AGG TGC TGG GTC CAA ATC ACC CCC ACA CTG TCA GCC CCG ACC TTC GGA GCG GTC ACG GCT CCT CTT CGG AGG GCC GTT GAC TAC TTA GCG GGA GGA GCT GCT CTC TGC TCC GCA CTA TAC GTC GGC GAC GCC TGC GGG	39 78 117 156 195 234 273
20	GCA GTG TTT CTG GTA GGC CAA ATG TTC ACC TAT AGG CCT CGC CAG CAT ACC ACA GTG CAG GAC TGC AAC TGT TCC ATT TAC AGT GGC CAT ATC ACC GGC CAC CGG ATG GCT TGG GAC ATG ATG ATG ATT AGC GC CAC CGG ATG GCT TGG GAC CAC AGG ATG GTT AGG ACC CTG ATG GCC CAG ATG GTA CGC ATG GCC CAG ATG CTC ACG ACC CTG ATG GCC CAG ATG CTT AGG GAC ATC CCC CAG GTG GTC ATA GAC ATC ATA GCC GGG GGC CAC TGG GGG GTC TTT GCC GCC GCA TAC TTT GCG TCC GCC GCA GTG GTC TAG GTA GTG CTG GTT CTG TTC CTG TTT GCG GGG GTC GAT GGC	312 351 390 429 468 507 546 576
25		
30	(2) INFORMATION FOR SEQ ID NO:46:  (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA4	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:	
35		

372577\_1

0		
5	GTT CCC TAC CGA AAC GCC TCT GGG GTT TAT C AAT GAT TGC CCA AAC TCT TCC ATA GTT TAC G AAC CTG ATC TTG CAT GCA CCT GGT TGC GTG C CCC ACG TTG TCA GCC CCG AAT CTC GGA GCC C CCT CTT CGG AGG GCC CTT GAC TAC CTC GGA GCG G CCC TC TGC TCC GCA CTA TAC GTC GGG ACC G CCC CAC TTT TT G GTA GCC CAA ATT TCA CCC G CCC CAG CAC ACT ACC GCA CAA ATC TTC ACC T CGC CAG CAA CCT ACC GTC CAA GAC TGC AAT TAC AGT GGC CAT ATC ACC GGC CAC CAA TG ATG ATG ATT GAT TCC CAC GCC TC ACT CAC CAT CAC CAC CAC ACC ACC ACC ACC	AG GCT GAT 78  CCT TGT GTC 117  CAA ATC ACC 156  GCA GCG GCT 195  SGA GGG GCT 195  SGA GGG GCT 234  SGC TGC GGG 273  ATA AGG CCT 312  GGC TCT ATT 351  GGA TGG GAC 390  CTG CTG ATG 429  ATC GAC ATC 468  SGC GCC GCA 507
10	GIC TIG TIT CIG TIT GCG GGG GTC GILL GCC	
15	(2) INFORMATION FOR SEQ ID NO:47:  (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pai (B) TYPE: nucleic acid (C) STRANDENNESS: single (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapien (C) INDIVIDUAL ISOLATE:  (xi) SEQUENCE DESCRIPTION: SEQ	e ns SA5
20	GTC CCC TAC CGA AAT GCC TCT GGG GTT TAT C	CAT GTC ACC 39

GTC CCC TAC CGA AAT GCC TCT GGG GTT TAT CAT GTC ACC AAT GAT TGC CCA AAC TCT TCC ATA GTC TAC GAG GCT GAT AAC CTG ATT CTG CAC GCA CCT GGT TGC GTG CCC TGT GTC AAG GAA GGT AAT GTC AGT AGG TGC TGG GTC CAA ATC ACC CCC ACA TTG TCA GCC CCG AAC CTC GGA GCG GTC ACG GCT CCT CTT CGG AGG GTC GTT GAC TAC TTA GCG GGA GGG GCT GCC CTC TGC TCC GCA CTA TAC GTC GGG GAC GCG TGC GGG GCA GTG TTC TTG GTA GGC CAA ATG TTC ACC TAT AGG CCT 25 CGC CAG CAT ACT ACG GTG CAG GAC TGC AAC TGT TCC ATT TAC AGC GGC CAT ATC ACC GGC CAC CGA ATG GCA TGG GAC ATG ATG ATG AAT TGG TCA CCT ACG ACA GCC TTG GTG ATG GCC CAG GTG CTA CGG ATT CCC CAA GTG GTC ATT GAC ATC ATT GCC GGG GGC CAC TGG GGG GTC TTG TTC GCC GTC GCA TAC TTC GCG TCA GCG GCT AAC TGG GCT AAG GTT GTG CTG GTC CTG TTT CTG TTT GCG GGG GTC GAT GGC 30

78

156

195

234

273

312

351

390 429

468

507

546

576

### (2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 576 base pairs
(B) TYPE: nucleic acid

٥		
	<pre>(C) STRANDEDNESS: single (D) TOPOLOGY: linear</pre>	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA6	
5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:	
10	GTT CCT TAC CGG AAT GCC TCT GGG GTG TAT CAT GTT ACC AAT GAT TGC CCA AAC TCT TCC ATA GTC TAT GAG GCT GAT GAC CTG ATC CTA CAC GCA CCT GGC TGC GTG CCC GCT GTC CGG AAG GAT AAT GTC ACT ACC GCA CCT GGC GTG GTC GTC GTC CCC ACA CTA TCA GCC CCG ASC CTC GGA GGG GTC ACG GCT CCT CTT CGG AGG GCC GTT ATA TAC TTG GCG GTC ACG GCT CCC CTG TGC TCC GCG TTA TAC GTC GGA GGG GTC ACG GCT CGC CTG TGC TCC GCG TTA TAC GTC GGA GGC GTC ACG GCT CGC CAG CAT GCT ACG GTA ACG CCAA ATG TTC ACC TAT AGG CCT CGC CAG CAT GCT ACG GTA CAG GAC TCC AAC TGC TCC ATT TAC AGG GC TA TC ACT GCC CAC CGG ATG GCA TGC ACG ATG ATG ATG AAT TCG ACT GCC CAC GGG GTG GCA TGC ACG ATG CTC CGG GGC CAC TGG GGG GCC AATG CTC CCC CAG GTG GCA TCC CCC CAG CAT TCC ACC TCC CCC CAG GTG GCC TTG GTC ATT GCC GGG GCC CAC TGG GGG GTC TTG TTC GCC GCT TAC TTT CTC GCC TCG GGG GCT AAC TGG GTC TTG GTC GTC TAC TTT TTT CTG TTT CTG GGG GGG GTT GAT GAT GTC TTT TTT TTT TTT TTT TTT GCG GGG GGT GAT GAT GAT GAT GTC TTT TTT TTT TTT TTT CTTT GCG GGG GGT GAT GAT GAT GAT GTC TTG GTC TTT TTT TTT CTTT CGG GGG GGT GAT GAT GAT GAT GAT GTC TTT TTT TTT TTT TTT TTT GCG GGG GTC TTG GTC TTG GTC TTT GTT CTTT TTT TTT GCG GGG GGT TTG TAT GAT GCC	39 78 117 156 195 234 273 312 351 390 429 468 507 546
	GTC TTG TTT CTG TTT GCG GGG GTT GAT GCC	576
	(2) INFORMATION FOR SEQ ID NO:49:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 576 base pairs (B) TYPE: nucleic acid (C) STRANDEDMESS: single (D) TOPOLOGY: linear	
	<ul><li>(vi) ORIGINAL SOURCE:</li><li>(A) ORGANISM: homosapiens</li><li>(C) INDIVIDUAL ISOLATE: SA7</li></ul>	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:	
30	GTC CCC TAC CGA AAT GCC TCC GGG GTT TAT CAT GTC ACC AAT GAT GAT GC CGA ACA TCT TCC ATA GTC TAT GAG GCT GAC AAC CTG ATC CTG GCC GCA CCT GGT TCC GTG CCC TGT GTC AGA CAA AAT AAT GTC AGT AGG TG GTC GGA GCT CCC ACA TTG TCA GCC CCG AAC CTC GGA GGG GTC ACG GCT CTT CTT CGG AGG GCT AGC CTC TCT TCG AGG GCC ACC CTC TG TCT CGC AGC CTC AGC CTC TG TC TC GCC AGA CTC TC AGC AGC CTC AGC AGC ACC ACC ACC ACC ACC ACC ACC AC	39 78 117 156 195 234 273 312 351 390 429 468 507
35	ATT GCC GGG GGC CAC 100 000 GTC TTG TTC 000 000	50,

0		
	TAT TTC GCG TCA GCG GCT AAC TGG GCT AAG GTT GTG CTG GTC TTG TTT CTG TTT GCG GGG GTC GAT GCC	546 576
	(2) INFORMATION FOR SEQ ID NO:50:	
5	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	<pre>(vi) ORIGINAL SOURCE:     (A) ORGANISM: homosapiens     (C) INDIVIDUAL ISOLATE: SA13</pre>	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:	
15	GTT CCC TAC CGA AAT GCC TCT GGG GTT TAT CAT GTC ACC AAT GAT TAC CCA AAC TCT TCC ATC GTG GTG CCT GAT GAT GAT TAC CAC GCA CCT GAT GAT GAT GAT TTA CAC GCA CCT GGT TCC GGG GTC CAG ATC ACC CCC ACA CTC TCA GCC CCG ACC CTC GGA GCG GTC CAG ATC ACC CCT CTT CGG AGG GCC GTA GAT TAC GCC CTT TGC TCC GCA GCC TTA TAC GTC GCA GCG GTG GCG GCG GTC CCT CTT TTG TTC GCA GCC CTT ATA CCT CCC CCT CTT TGC TCC GCC GCA ACC CTC GCA GCC TTA GCC CCT CCC CCG GCA GCG GTG TCC CCT CCC CCG GCA ATC GTC CCC CCG CAT AAT GTT GCC GCC CAC ATC ACC GCC CAC ATC ACC GCC CAC ATC ACC ACC ACC ACC ACC ACC ACC A	39 78 117 156 195 234 273 312 351 390 429 468 507 546 576
	(2) INFORMATION FOR SEQ ID NO:51:	
25	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 576 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
30	(Vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: HK2	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:	
35	CTT ACC TAC GGC AAC TCC AGT GGG CTA TAC CAT CTC ACA AAT GAT TGC CCC AAC TCC AGC ATC GTG CTG GAG GCG GAT GCT ATG ATC TTG CAT TTG CCT GGA TGC TTG CCT TGT GTG AGG GTC GAT GAT CGG TCC ACC TGT TGG CAT GCT GTG ACC	39 78 117 156

372577\_1

.

15

CCC ACC CTG GCC ATA CCA AAT GCT TCC ACG CCC GCA ACG 195 GGA TTC CGC AGG CAT GTG GAT CTT CTT GCG GGC GCC GCA 234 GTG GTT TGC TCA TCC CTG TAC ATC GGG GAC CTG TGT GGC 273 ACC TTT CAG CCC 312 TCT CTC TTT TTG GCG GGA CAA CTA TTC CGC CGT CAT TGG ACT GTG CAA GAC TGC AAC TGC TCC ATC 351 TAT ACA GGC CAC GTC ACC GGC CAC AGG ATG GCT TGG GAC 390 ATG ATG AAC TGG TCA CCC ACA ACC ACT CTG GTC CTA 429 ATC TTG AGG GTA CCT GAG ATT TGT GCG AGT GTG 468 ATA TTT GGT GGC CAT TGG GGG ATA CTA CTA GCC GTT GCC 507 TAC TTT GGC ATG GCT GGC AAC TGG CTA AAA GTT CTG GCT 546 576 GTT CTG TTC CTA TTT GCA GGG GTT GAA GCA

## (2) INFORMATION FOR SEQ ID NO:52:

10 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE: (A) ORGANISM: h

(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: DK7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Tyr Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asp 15 10 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu 20 30 20 25 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Val Ser 45 35 40 Arg Cys Trp Val Ala Met Thr Pro Thr Val Ala Thr Arg Asp Gly 60 5.0 55 Lys Leu Pro Thr Ala Gln Leu Arg Arg His Ile Asp Leu Leu Val 75 70 65 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys 25 90 85 80 Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg 100 95 Arg His Trp Thr Thr Gln Gly Cys Asn Cys Ser Ile Tyr Pro Gly 120 110 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 130 135 30 Ser Pro Thr Thr Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro 150 140 145 Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu 160 155 Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val 180 175

Leu Val Val Leu Leu Phe Ala Gly Val Asp Ala 185

### INFORMATION FOR SEQ ID NO:53: (2)

SEQUENCE CHARACTERISTICS: (i)

192 amino acids (A) LENGTH:

amino acid TYPE:

STRANDEDNESS: unknown (C) TOPOLOGY: unknown

(D)

### (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

INDIVIDUAL ISOLATE: (C)

SEQUENCE DESCRIPTION: SEQ ID NO:53: (xi)

Tyr Gln Val Arg Asn Ser Ser Gly Leu Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu 20 25 His Ser Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser 35 Lys Cys Trp Val Ala Val Ala Pro Thr Val Ala Thr Arg Asp Gly Lys Leu Pro Ala Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val 65 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys 85 90 80 Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg 105 95 100 Arq His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 120 110 115 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 130 135 125 Ser Pro Thr Ala Ala Leu Val Met Ala Gln Leu Leu Arg Ile Pro 145 150 140 Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val 170 175 Val Val Val Leu Leu Phe Thr Gly Val Asp Ala

30

5

10

15

20

25

185 INFORMATION FOR SEO ID NO:54: (2)

> SEQUENCE CHARACTERISTICS: (i) LENGTH: 192 amino acids (A)

> > TYPE: amino acid (B)

190

STRANDEDNESS: unknown (C)

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: DR1
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

His Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asn 10 15

Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu 25

His Ala Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser 35

Arg Cys Trp Val Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly 60

- 95 100 105

  15 Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly
  110 115 120

  His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp
  - His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asi Trp 125 130 135

    Ser Pro Thr Thr Ala Leu Val Met Ala Gln Leu Leu Arg Ile Pro 140 145 150
    - Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu 155 160 165 Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val
    - 170 175 Val Val Val Leu Leu Phe Ala Gly Val Asp Ala 185 190
      - (2) INFORMATION FOR SEQ ID NO:55:
- 25 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
  - (vi) ORIGINAL SOURCE:
    - (A) ORGANISM: homosapiens
    - (C) INDIVIDUAL ISOLATE: DF
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:
  - His Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asp 5 10 15

30

Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu 25 20 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Thr Ser 40 45 35 Arg Cys Trp Val Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly 50 55 60 Lys Leu Pro Thr Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val 70 75 65 5 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys 90 85 80 Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg 105 His His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 115 110 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 135 130 125 10 Ser Pro Thr Thr Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro 145 140 Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val 175 170 Leu Val Val Leu Leu Phe Ala Gly Val Asp Ala 15 185

### INFORMATION FOR SEO ID NO:56: (2)

SEQUENCE CHARACTERISTICS: (i)

LENGTH: 192 amino acids (A)

amino acid (B) TYPE:

STRANDEDNESS: (D)

TOPOLOGY: unknown

ORIGINAL SOURCE: (vi) (A)

ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE:

SEQUENCE DESCRIPTION: SEQ ID NO:56: 25 (xi)

> Tyr Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asp 15 10 Cys Pro Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Ala Ile Leu 30 20 His Ala Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Thr Ser 40 45 35 Arg Cys Trp Val Ala Met Thr Pro Thr Val Ala Thr Arg Asp Gly 60 50 Lys Leu Pro Ala Thr Gln Leu Arg Arg Tyr Ile Asp Leu Leu Val 70 65 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys 90 85 80

35

30

Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg Arg Leu Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His Arq Met Ala Trp Asp Met Met Asn Trp Ser Pro Thr Thr Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Phe Ala Gly Val Asp Ala 

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: S18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Tyr Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Thr Ile Leu His Ser Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser Trp Val Pro Val Ala Pro Thr Val Ala Thr Arg Asp Gly Lys Leu Pro Ala Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Thr Ala Leu Val Ile Ala Gln Leu Leu Arg Val Pro Gln Ala Val Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu 

- Ala Gly Ile Ala Tyr Phe Ser Met Ala Gly Asn Trp Ala Lys Val 170 175 Leu Leu Val Leu Leu Phe Ala Gly Val Asp Ala 185
  - (2) INFORMATION FOR SEO ID NO:58:
- 5 (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
    (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
  - (vi) ORIGINAL SOURCE:
    - (A) ORGANISM: homosapiens
      - (C) INDIVIDUAL ISOLATE: SI
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Tyr Gln Val Arg Asn Ser Ser Gly Leu Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Ala Ile Leu 15 20 His Ser Pro Gly Cys Val Pro Cys Val Arg Glu Asp Gly Ala Pro Lys Cys Trp Val Ala Val Ala Pro Thr Val Ala Thr Arg Asp Gly Lys Leu Pro Ala Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val 65 70 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys 20 85 90 80 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 Arq His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 120 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Asn Trp 125 130 135 Ser Pro Thr Thr Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro 25 145 140 Gln Ala Val Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu

185
(2) INFORMATION FOR SEQ ID NO:59:

155

170

(i) SEQUENCE CHARACTERISTICS:

Leu Ile Val Leu Leu Phe Ser Gly Val Asp Ala

(A) LENGTH: 192 amino acids

Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val

160

175

(B) TYPE: amino acid

35

30

- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: US11
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

20 Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu
155 160
Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val
170 175 180
Leu Val Val Leu Leu Phe Ala Gly Val Asp Ala
185 190

- 25 (2) INFORMATION FOR SEQ ID NO:60:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 192 amino acids
      - (B) TYPE: amino acid
      - (C) STRANDEDNESS: unknown
      - (D) TOPOLOGY: unknown
- - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: D1
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

c

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Met Ile Met 20 25 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asp Asn Ser Ser 35 40 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Gly 55 5 Asn Val Pro Thr Thr Ala Ile Arg Arg His Val Asp Leu Leu Val 70 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 90 80 8.5 Gly Ser Val Phe Leu Ile Ser Gln Leu Phe Thr Leu Ser Pro Arg 100 105 95 Arq His Glu Thr Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly 120 115 110 10 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 140 145 Gln Ala Val Met Asp Met Val Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 15 170 175 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 185 190

## (2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens
(C) INDIVIDUAL ISOLATE: D3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

35

20

25

15

25

30

Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 90 85 80 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 100 95 Arg His Glu Thr Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 120 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 135 130 125 5 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 145 140 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 170 175 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 185 10

### (2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: 1

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Val Asp Val Ile Met 30 2.0 25 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn His Ser 35 40 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala 55 Ser Ile Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 75 70 65 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 90 80 85 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 100 105 95 Arg His Glu Thr Ala Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 120 110 115 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Asn Trp 125 130 Ser Pro Thr Thr Ala Leu Val Leu Ser Gln Leu Leu Arg Ile Pro 145 150 140

° Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu
155
Ala Gly Leu Ala Tyr Tyr Ser Met Ala Gly Asn Trp Ala Lys Val
170
Leu Ile Val Leu Leu Phe Ala Gly Val Asp Gly
185

## 5 (2) INFORMATION FOR SEQ ID NO:63:

- (i) SEOUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
  - (vi) ORIGINAL SOURCE:
    - (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:
- Tyr Glu Val Arg Asn Val Ser Gly Ile Tyr His Val Thr Asn Asp 15 10 Cys Ser Asn Ser Ser Val Val Tyr Glu Thr Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser 35 40 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Val 55 Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 20 70 75 65 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 90 80 85 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 105 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Leu Tyr Pro Gly 110 115 120 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 25 125 130 135
- Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 140 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu
- 155 160 165
  Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
  170 175 180
- Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 185 190
  - (2) INFORMATION FOR SEO ID NO:64:
  - (i) SEQUENCE CHARACTERISTICS:

35

LENGTH: 192 amino acids

```
(B)
                    TYPE: amino acid
                    STRANDEDNESS: unknown
               (C)
               (D)
                    TOPOLOGY:
                              unknown
               ORIGINAL SOURCE:
                    ORGANISM:
                               homosapiens
               (A)
               (C)
                    INDIVIDUAL ISOLATE:
               SEQUENCE DESCRIPTION: SEQ ID NO:64:
His Glu Val His Asn Val Ser Gly Ile Tyr His Val Thr Asn Asp
                                      10
                                                          15
Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met
                                                           30
                 20
                                      25
His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser
                                      40
Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala
                                      55
Ser Ile Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val
                                      70
                 65
Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys
                                      85
                                                           90
                 80
Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg
                                     100
                 95
Arq His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly
                                     115
                                                          120
                110
His Val Ser Gly His Arg Met Ala Trp Asp Met Met Asn Trp
                                     130
                                                          135
                125
Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Leu Pro
                                                          150
                140
                                     145
Gln Ala Val Met Asp Met Val Ala Gly Ala His Trp Gly Val Leu
                155
                                     160
Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
                                     175
                                                          180
Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly
```

190

25

0

5

10

15

20

185 (2) INFORMATION FOR SEO ID NO:65:

(A)

(vi)

(xi)

SEQUENCE CHARACTERISTICS: (i)

LENGTH: 192 amino acids (A) TYPE: amino acid (B)

(C) STRANDEDNESS: unknown

TOPOLOGY: unknown (D)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens INDIVIDUAL ISOLATE: HK5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

372577 1

35

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp Cys Ser Asn Leu Ser Ile Val Tyr Glu Thr Thr Asp Met Ile Met 25 20 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser 35 40 Arg Cys Trp Val Ala Leu Ala Pro Thr Leu Ala Ala Arg Asn Ala 50 Ser Val Pro Thr Thr Ala Ile Arg Arg His Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 85 90 80 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 120 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 140 145 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 170 180 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 185

### (2) INFORMATION FOR SEQ ID NO:66:

20 (i)

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

25 (vi)

ORIGINAL SOURCE:
(A) ORGANISM: homosapiens
(C) INDIVIDUAL ISOLATE: HK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Tyr Glu Val Arg Asn Val Ser Gly Ile Tyr His Val Thr Asn Asp 5

Cys Ser Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Met Ile Met 20

His Thr Pro Gly Cys Met Pro Cys Val Arg Glu Asn Asn Ser Ser 35

Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Val 50

Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 65

70

35

30

٥

10

Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 80 90 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 115 110 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 145 140 Gln Ala Ile Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 180 170 175 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 185

### (2) INFORMATION FOR SEO ID NO:67:

SEQUENCE CHARACTERISTICS: (i) LENGTH: 192 amino acids

amino acid (B) TYPE:

(C)

STRANDEDNESS: unknown TOPOLOGY: unknown (D)

### (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

INDIVIDUAL ISOLATE: IND5 (C)

SEQUENCE DESCRIPTION: SEQ ID NO:67: (xi)

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met 20 25 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ser Ser 35 40 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala 50 55 Ser Val Ser Thr Thr Thr Ile Arg His His Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 85 90 80 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 100 95 Arq His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 145 150 140

35

0

5

10

15

20

2.5

Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Ile Leu
155
Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
170
Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly
185
185
190

5

10

0

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unk

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens
(C) INDIVIDUAL ISOLATE: INDE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

15 Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp 10 15 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Phe Ser 35 40 Ser Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala 55 20 Ser Val Pro Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 70 75 65 Gly Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 85 80 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 100 95 105 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 25 110 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 135 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 140 145 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Ile Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 30 175 170 180 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly

190

185
(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

5 (vi) ORIGINAL SOURCE:
(A) ORGANISM: homosapiens
(C) INDIVIDUAL ISOLATE: P10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Tvr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met 10 25 20 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser 40 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ser 55 50 Ser Val Pro Thr Thr Ala Ile Arg Arg His Val Asp Leu Leu Val 70 15 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 85 90 80 Gly Ser Val Leu Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 105 95 Arg His Trp Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 115 120 110 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 135 20 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 140 145 Gln Ala Ile Leu Asp Val Val Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 170 175 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 25 185

### (2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens
(C) INDIVIDUAL ISOLATE: S9

35

### (xi) SEOUENCE DESCRIPTION: SEO ID NO:70:

Tyr Glu Val Arg Asn Val Ser Gly Ala Tyr His Val Thr Asn Asp 15 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Val Ile Met 20 25 His Thr Pro Gly Cys Val Pro Cys Val Gln Glu Gly Asn Ser Ser 5 35 40 Gln Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala 50 55 Thr Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 65 Gly Ala Ala Val Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 90 80 85 Gly Ser Val Phe Leu Ile Ser Gln Leu Phe Thr Ile Ser Pro Arq 10 95 100 105 Arg His Glu Thr Val Gln Asn Cys Asn Cys Ser Ile Tyr Pro Gly 120 110 115 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 135 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 145 140 15 Gln Ala Val Met Asp Met Val Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 170 175 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 185

20 (2) INFORMATION FOR SEO ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

25

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens(C) INDIVIDUAL ISOLATE: S45

(xi) SEOUENCE DESCRIPTION: SEO ID NO:71:

30 Tyr Glu Val Arg Asn Val Ser Gly Ala Tyr His Val Thr Asn Asp 5 10 15

Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Val Asp Val Ile Leu 20 25 30

His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser 40 45

Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ser 50 60

Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 65 70 Gly Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 85 90 80 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 130 125 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 140 145 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 170 175 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 190 185

## (2) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
    (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: SA10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Tyr Glu Val Arg Asn Val Ser Gly Met Tyr His Val Thr Asn Asp 1.5 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met 20 25 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser 35 40 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ser 50 55 Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 85 90 80 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 Arg Tyr Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 115 120 Arg Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 135

30

5

10

15

20

Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro
140
Gln Ala Ile Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu
155
Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
170
Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly
185
185

5

10

15

20

25

## (2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: SW2
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Tyr Glu Val Arq Asn Val Ser Gly Val Tyr His Val Thr Asn Asp 10 15 Cvs Ser Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Met Ile Met 25 30 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Ala Asn Ser Ser 35 40 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Thr 55 50 Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 65 70 75 Gly Ala Ala Ala Phe Cys Ser Val Met Tyr Val Gly Asp Leu Cys 80 85 90 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 105 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 135 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro

145

Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu
155 160 165
Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val
170 175 180

Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 185 190

(2) INFORMATION FOR SEQ ID NO:74:

140

35

372577\_1

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

5 (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: T3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr Tyr Val Thr Asn Asp 15 10 10 Cys Ser Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Met Ile Met 25 20 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Ser Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala 50 55 Ser Val Pro Thr Lys Thr Ile Arg Arg His Val Asp Leu Leu Val 15 70 65 Gly Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 90 80 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 105 Arq His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 115 120 110 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Asn Trp 20 125 130 135 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 140 145 150 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 170 175 180 25 Leu Ile Val Leu Leu Phe Ala Gly Val Asp Gly

# 185 (2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

### (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: T10

35

30

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Tyr Glu Val Arg Asn Val Ser Gly Met Tyr His Val Thr Asn Asp 10 Cys Ser Asn Ser Ser Ile Val Phe Glu Ala Ala Asp Leu Ile Met 25 5 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ser Ser 35 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Thr 55 60 Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 65 70 75 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 80 85 90 10 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg 100 105 95 Arq His Glu Thr Leu Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 120 His Leu Ser Gly His Arq Met Ala Trp Asp Met Met Met Asn Trp 125 130 135 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 15 145 140 Gln Ala Val Met Asp Met Val Thr Gly Ala His Trp Gly Val Leu 160 165 155 Ala Gly Leu Ala Tyr Tyr Ser Met Ala Gly Asn Trp Ala Lys Val 170 175 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly 185

20

25

٥

### (2) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: US

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Tyr Glu Val Arg Asn Val Ser Gly Met Tyr His Val Thr Asn Asp 5 10 15 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met 20 25 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser 40 45

35

Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val 70 75 65 Gly Ala Ala Thr Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys 80 85 90 Gly Ser Val Phe Leu Ile Ser Gln Leu Phe Thr Phe Ser Pro Arg 95 100 105 5 Gln His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 115 120 110 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro 140 145 150 Gln Ala Val Met Asp Met Val Ala Gly Ala His Trp Gly Val Leu 155 160 10 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val 175 170 Leu Ile Val Leu Leu Phe Ala Gly Val Asp Gly 185

### (2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
  - 192 amino acids (A) LENGTH:
  - (B) TYPE: amino acid STRANDEDNESS:
  - (C) unknown unknown
  - (D) TOPOLOGY:
  - (vi) ORIGINAL SOURCE:

125

- (A) ORGANISM: homosapiens 20
  - (C) INDIVIDUAL ISOLATE:
  - (xi) SECUENCE DESCRIPTION: SEO ID NO:77:

Ala Gln Val Arq Asn Thr Ser Arg Gly Tyr Met Val Thr Asn Asp Cys Ser Asn Glu Ser Ile Thr Trp Gln Leu Gln Ala Ala Val Leu 25 2.0 His Val Pro Gly Cys Ile Pro Cys Glu Arg Leu Gly Asn Thr Ser Arg Cys Trp Ile Pro Val Thr Pro Asn Val Ala Val Arg Gln Pro Gly Ala Leu Thr Gln Gly Leu Arg Thr His Ile Asp Met Val Val 75 65 70 Met Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys 30 85 90 80 Gly Gly Val Met Leu Ala Ala Gln Met Phe Ile Val Ser Pro Arg 95 100 105 Arg His Trp Phe Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 120 Thr Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp

130

135

35

 Ser Pro
 Thr
 Ala
 Thr
 Met
 Ile Leu Ala
 Tyr
 Ala Met
 Arg Val
 Pro

 Glu
 Val
 Ile Ile Asp
 Ile Gly Gly
 Ala His
 Trp Gly
 Val
 Met

 Ibe
 1.55
 160
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165
 165

## (2) INFORMATION FOR SEO ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: T4

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Ala Gln Val Lys Asn Thr Thr Asn Ser Tyr Met Val Thr Asn Asp 10 15 Cys Ser Asn Asp Ser Ile Thr Trp Gln Leu Gln Ala Ala Val Leu 30 20 His Val Pro Gly Cys Val Pro Cys Glu Lys Thr Gly Asn Thr Ser 35 40 Arg Cys Trp Ile Pro Val Ser Pro Asn Val Ala Val Arg Gln Pro 50 55 60 Gly Ala Leu Thr Gln Gly Leu Arg Thr His Ile Asp Met Val Val 65 70 75 Met Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys 80 85 90 Gly Gly Val Met Leu Ala Ala Gln Met Phe Ile Val Ser Pro Gln 95 100 His His Trp Phe Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 120 Thr Ile Thr Gly His Arq Met Ala Trp Asp Met Met Met Asn Trp 125 Ser Pro Thr Ala Thr Met Ile Leu Ala Tyr Ala Met Arg Val Pro 145 150 140 Glu Val Ile Leu Asp Ile Val Ser Gly Ala His Trp Gly Val Met 155 160 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val 170 175 180 Val Val Ile Leu Leu Leu Ala Ala Gly Val Asp Ala 185 190

30

0

5

10

20

0

5

(2) INFORMATION FOR SEQ ID NO:79:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
    - STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- ORIGINAL SOURCE: (vi)
  - homosapiens (A) ORGANISM:
  - INDIVIDUAL ISOLATE:
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:
- Ala Glu Val Lys Asn Thr Ser Thr Ser Tyr Met Val Thr Asn Asp 10 10 Cys Ser Asn Asp Ser Ile Thr Trp Gln Leu Gln Ala Ala Val Leu 20 25 His Val Pro Gly Cys Val Pro Cys Glu Arg Val Gly Asn Ala Ser 35 40 Arg Cys Trp Ile Pro Val Ser Pro Asn Val Ala Val Gln Arg Pro 50 55
- 15 Gly Ala Leu Thr Gln Gly Leu Arg Thr His Ile Asp Met Val Val 65 70 Met Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys
- 85 90 80 Gly Gly Val Met Leu Ala Ala Gln Met Phe Ile Ile Ser Pro Gln 95 100 105
- His His Trp Phe Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 120 20 Thr Ile Thr Gly His Arq Met Ala Trp Asp Met Met Met Asn Trp
  - 125 130 135 Ser Pro Thr Thr Thr Met Ile Leu Ala Tyr Ala Met Arg Val Pro 140 145
- Glu Val Ile Ile Asp Ile Ile Ser Gly Ala His Trp Gly Val Met 155 160 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val
- 25 170 175 Val Val Ile Leu Leu Thr Ala Gly Val Asp Ala 185 190
  - (2) INFORMATION FOR SEO ID NO:80:
- (i) SEOUENCE CHARACTERISTICS: 30
  - (A) LENGTH: 192 amino acids TYPE: amino acid
  - (B)
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
  - ORIGINAL SOURCE: (vi)
- (A) ORGANISM: homosapiens 35

### INDIVIDUAL ISOLATE: US10

### (xi) SEQUENCE DESCRIPTION: SEO ID NO:80:

Val Gln Val Lys Asn Thr Ser Thr Ser Tyr Met Val Thr Asn Asp 10 Cys Ser Asn Asp Ser Ile Thr Trp Gln Leu Glu Ala Ala Val Leu 5 His Val Pro Gly Cys Val Pro Cys Glu Lys Val Gly Asn Thr Ser Arg Cvs Trp Ile Pro Val Ser Pro Asn Val Ala Val Gln Arg Pro 50 55 60 Gly Ala Leu Thr Gln Gly Leu Arg Thr His Ile Asp Met Val Val 65 70 Met Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Phe Cys 10 80 85 90 Gly Gly Met Met Leu Ala Ala Gln Met Phe Ile Val Ser Pro Arq 95 100 105 His His Ser Phe Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 Thr Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 135 15 Ser Pro Thr Ala Thr Leu Ile Leu Ala Tyr Val Met Arg Val Pro 140 145 Glu Val Ile Ile Asp Ile Ile Ser Gly Ala His Trp Gly Val Leu 155 160 165 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val 170 175 180 Val Val Ile Leu Leu Leu Ala Ala Gly Val Asp Ala 185

### (2) INFORMATION FOR SEO ID NO:81:

(i) SEOUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids (B) TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown (D)

ORIGINAL SOURCE: (vi)

> (A) ORGANISM: homosapiens

INDIVIDUAL ISOLATE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Val Glu Val Arg Asn Ile Ser Ser Ser Tyr Tyr Ala Thr Asn Asp 10 Cys Ser Asn Asn Ser Ile Thr Trp Gln Leu Thr Asp Ala Val Leu 20 25 His Leu Pro Gly Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu 35 40

35

20

25

Arg Cys Trp Ile Gln Val Thr Pro Asn Val Ala Val Lys His Arg 50 55 Gly Ala Leu Thr His Asn Leu Arg Thr His Val Asp Val Ile Val 65 70 Met Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Val Cys 85 80 Gly Ala Val Met Ile Val Ser Gln Ala Leu Ile Ile Ser Pro Glu 5 95 Arg His Asn Phe Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly 110 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Leu Asn Trp 135 125 130 Ser Pro Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro 140 145 10 Glu Leu Ala Leu Gln Val Val Phe Gly Gly His Trp Gly Val Val 155 160 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val 170 175 Ile Ala Ile Leu Leu Val Ala Gly Val Asp Ala 185 190

## 15 (2) INFORMATION FOR SEO ID NO:82:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- 20
   (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: I

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Val Glu Val Arg Asn Thr Ser Ser Ser Tyr Tyr Ala Thr Asn Asp 25 Cys Ser Asn Asn Ser Ile Thr Trp Gln Leu Thr Asn Ala Val Leu 20 His Leu Pro Gly Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu 35 45 His Cys Trp Ile Gln Val Thr Pro Asn Val Ala Val Lys His Arg 50 55 60 Gly Ala Leu Thr His Asn Leu Arg Ala His Ile Asp Met Ile Val 30 65 70 Met Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Val Cys 80 85 90 Gly Ala Val Met Ile Val Ser Gln Ala Phe Ile Val Ser Pro Glu 95 100 His His His Phe Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly 110 115 120

His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Leu Asn Trp 125
Ser Pro Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro 140
Glu Leu Val Leu Glu Val Val Phe Gly Gly His Trp Gly Val Val Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val Ile Ala Ile Leu Leu Val Ala Gly Val Asp Ala 185
180

### (2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid (C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Val Glu Val Arg Asn Ile Ser Ser Ser Tyr Tyr Ala Thr Asn Asp 10 15 Cvs Ser Asn Ser Ser Ile Thr Trp Gln Leu Thr Asn Ala Val Leu 25 3.0 20 His Leu Pro Gly Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu 40 His Cys Trp Ile Gln Val Thr Pro Asn Val Ala Val Lys His Arg 55 Gly Ala Leu Thr His Asn Leu Arg Ala His Val Asp Met Ile Val 70 65 Met Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Met Cys 25 85 80 Glv Ala Val Met Ile Val Ser Gln Ala Phe Ile Ile Ser Pro Glu 95 100 Arg His Asn Phe Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly 115 110 120 Arg Ile Thr Gly His Arg Met Ala Trp Asp Met Met Leu Asn Trp 125 130 135 Ser Pro Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro 30 140 145 150 Glu Leu Val Leu Glu Val Val Phe Gly Gly His Trp Gly Val Val 155 160 165 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val 170 180 Ile Ala Ile Leu Leu Val Ala Gly Val Asp Ala 185 190

35

5

10

5

15

20

2.5

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: 7

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Val Glu Val Arg Asn Thr Ser Phe Ser Tyr Tyr Ala Thr Asn Asp 10 Cys Ser Asn Asn Ser Ile Thr Trp Gln Leu Thr Asn Ala Val Leu 20 25 His Leu Pro Gly Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu 35 40 Arg Cys Trp Ile Gln Val Thr Pro Asn Val Ala Val Lys His Arg 50 Gly Ala Leu Thr His Asn Leu Arg Thr His Val Asp Val Ile Val 65 70 Met Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Val Cys 80 85 90 Gly Ala Val Met Ile Ala Ser Gln Ala Phe Ile Ile Ser Pro Glu 95 100 105 Arg His Asn Phe Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly 110 115 120 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Leu Asn Trp 125 130 Ser Pro Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro 140 145 Glu Leu Val Leu Glu Val Val Phe Gly Gly His Trp Gly Val Val

Giu Leu Val Leu Glu Val Val Phe Gly Gly His Trp Gly Val Val
155
160
Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val
170
175
180
Ile Ala Ile Leu Leu Val Ala Gly Val Asp Ala
185
190

30 (2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: S83

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Val Glu Val Lys Asp Thr Gly Asp Ser Tyr Met Pro Thr Asn Asp 5 10 15 Cys Ser Asn Ser Ser Ile Val Trp Gln Leu Glu Gly Ala Val Leu 30 His Thr Pro Gly Cys Val Pro Cys Glu Arg Thr Ala Asn Val Ser 35 Arg Cys Trp Val Pro Val Ala Pro Asn Leu Ala Ile Ser Gln Pro 50 10 Gly Ala Leu Thr Lys Gly Leu Arg Ala His Ile Asp Ile Ile Val 65 70 Met Ser Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Val Cys 80 85 Gly Ala Leu Met Leu Ala Ala Gln Val Val Val Ser Pro Gln 95 100 105 His His Thr Phe Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly 110 115 120 15 Arg Ile Thr Gly His Arg Met Ala Trp Asp Met Met Asn Trp 125 130 Ser Pro Thr Thr Met Leu Leu Ala Tyr Leu Val Arg Ile Pro 140 145 150 Glu Val Ile Leu Asp Ile Val Thr Gly Gly His Trp Gly Val Met 155 160 165 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ser Trp Ala Lys Val 170 175 180 20 Ile Val Ile Leu Leu Thr Ala Gly Val Glu Ala 185

### (2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: DK1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

35

30

2.5

His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser 35 Thr Cys Trp Thr Ser Val Thr Pro Thr Val Ala Val Arg Tyr Val 50 55 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val 65 70 Gly Ala Ala Thr Met Cys Ser Ala Leu Tyr Val Gly Asp Val Cys 5 80 90 Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg 95 100 105 Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly 110 115 120 His Leu Ser Gly His Arg Met Ala Trp Asp Met Met Asn Trp 125 130 135 10 Ser Pro Ala Val Gly Met Val Val Ala His Val Leu Arg Leu Pro 140 145 Gln Thr Leu Phe Asp Ile Ile Ala Gly Ala His Trp Gly Ile Met 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val 170 175 Ala Ile Ile Met Val Met Phe Ser Gly Val Asp Ala 185 15

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: HK10
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Leu Glu Trp Arg Asn Val Ser Gly Leu Tyr Val Leu Thr Asn Asp 15 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Val Ile Leu 20 25 His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser 35 40 Thr Cys Trp Thr Ser Val Thr Pro Thr Val Ala Val Arg Tyr Val 50 55 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val 65 70 Gly Ala Ala Thr Met Cys Ser Ala Leu Tyr Val Gly Asp Met Cys 80 85 90 Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg 100 105

35

30

Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly 110 115 120 His Leu Ser Gly His Arg Met Ala Trp Asp Met Met Asn Trp 125 130 135 Ser Pro Ala Val Gly Met Val Val Ala His Val Leu Arg Leu Pro 140 145 Gln Thr Leu Phe Asp Ile Ile Ala Gly Ala His Trp Gly Ile Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val 170 175 180 Ala Ile Ile Met Val Met Phe Ser Gly Val Asp Ala 185

# (2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE: (A) ORGANISM: h

(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Leu Glu Trp Arg Asn Thr Ser Gly Leu Tyr Val Leu Thr Asn Asp 10 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Val Ile Leu 20 25 His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser 35 40 Thr Cys Trp Thr Pro Val Thr Pro Thr Val Ala Val Arg Tyr Val 50 55 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val 65 70 Gly Ala Ala Thr Met Cys Ser Ala Leu Tyr Val Gly Asp Met Cys 85 90 Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg 95 105 Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly 110 115 120 His Leu Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 130 135 Ser Pro Ala Val Gly Met Val Val Ala His Val Leu Arg Leu Pro 140 145 150 Gln Thr Val Phe Asp Ile Ile Ala Gly Ala His Trp Gly Ile Leu 155 160 Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val 170 175 180

30

5

10

15

20

Ala Ile Ile Met Val Met Phe Ser Gly Val Asp Ala 185

## (2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

# (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: S52

10

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Leu Glu Trp Arg Asn Thr Ser Gly Leu Tyr Val Leu Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Val Ile Leu 20 25 His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser 15 35 40 Met Cys Trp Thr Pro Val Thr Pro Thr Val Ala Val Arg Tyr Val 50 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val 70 75 Gly Ala Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Met Cys 80 85 90 Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg 20

95 100 105
Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly
110
His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp

125 130 135 Ser Pro Ala Val Gly Met Val Val Ala His Ile Leu Arg Leu Pro 140 145 150

Gln Thr Leu Phe Asp Ile Leu Ala Gly Ala His Trp Gly Ile Leu 155 155 160 160 165
Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val 170 175 180
Ala Ile Val Met Ile Met Phe Ser Gly Val Asp Ala

185 190

INFORMATION FOR SEQ ID NO:90:

SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

35

25

(i)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S54

(xi) SEQUENCE DESCRIPTION: SEO ID NO:90:

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:90:

5 Leu Glu Trp Arg Asn Thr Ser Gly Leu Tyr Ile Leu Thr Asn Asp 10 15 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Val Ile Leu 30 25 His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser 35 40 Thr Cys Trp Thr Pro Val Thr Pro Thr Val Ala Val Arg Tyr Val 10 50 55 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val 65 70 Gly Ala Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Met Cys 8.0 85

Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg 95 100 105 Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly

15 110 115 120

His Leu Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp
125 120

Ser Pro Ala Val Gly Met Val Val Ala His Ile Leu Arg Leu Pro
140 145 150

Gln Thr Leu Phe Asp Ile Leu Ala Gly Ala His Trp Gly Ile Leu
155 160 165

Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val
170 175 180
Ala Ile Ile Met Ile Met Phe Ser Gly Val Asp Ala
185 190

(2) INFORMATION FOR SEQ ID NO:91:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

C) INDIVIDUAL ISOLATE: Z4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Glu His Tyr Arg Asn Ala Ser Gly Ile Tyr His Ile Thr Asn Asp  $\phantom{0}5\phantom{0}$  10  $\phantom{0}15\phantom{0}$ 

35

Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp His His Ile Leu His Leu Pro Gly Cys Val Pro Cys Val Met Thr Gly Asn Thr Ser 35 40 Arg Cys Trp Thr Pro Val Thr Pro Thr Val Ala Val Ala His Pro 50 Gly Ala Pro Leu Glu Ser Phe Arg Arg His Val Asp Leu Met Val 65 70 5 Gly Ala Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys 80 85 90 Gly Gly Ala Phe Leu Met Gly Gln Met Ile Thr Phe Arg Pro Arg 95 100 105 Arg His Trp Thr Thr Gln Glu Cys Asn Cys Ser Ile Tyr Thr Gly 110 115 120 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Asn Trp 10 125 130 135 Ser Pro Thr Thr Leu Leu Leu Ala Gln Ile Met Arg Val Pro 140 145 150 Thr Ala Phe Leu Asp Met Val Ala Gly Gly His Trp Gly Val Leu 155 160 Ala Gly Leu Ala Tyr Phe Ser Met Gln Gly Asn Trp Ala Lys Val 170 175 Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala 15 185 190

#### (2) INFORMATION FOR SEQ ID NO:92:

(i)

SEQUENCE CHARACTERISTICS: (A) LENGTH: 192 amino acids

(B) TYPE: amino acid STRANDEDNESS:

unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE: (A) ORGANISM:

homosapiens (C) INDIVIDUAL ISOLATE:

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

Val His Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp 10 15 Cys Pro Asn Thr Ser Ile Val Tyr Glu Thr Glu His His Ile Met His Leu Pro Gly Cys Val Pro Cys Val Arg Thr Glu Asn Thr Ser 35 Arg Cys Trp Val Pro Leu Thr Pro Thr Val Ala Ala Pro Tyr Pro 50 55 Asn Ala Pro Leu Glu Ser Met Arg Arg His Val Asp Leu Met Val 65 70 Gly Ala Ala Thr Met Cys Ser Ala Phe Tyr Ile Gly Asp Leu Cys 80 85

35

30

Gly Gly Val Phe Leu Val Gly Gln Leu Phe Asp Phe Arg Pro Arg Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Ser Ala Leu Ile Met Ala Gln Ile Leu Arg Ile Pro Ser Ile Leu Gly Asp Leu Leu Thr Gly Gly His Trp Gly Val Leu Ala Gly Leu Ala Phe Phe Ser Met Gln Ser Asn Trp Ala Lys Val Ile Leu Val Leu Phe Leu Phe Ala Gly Val Glu Gly 

10
(2) INFORMATION FOR SEQ ID NO:93:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: Z
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:
- Val Asn Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Glu His Gln Ile Leu His Leu Pro Gly Cys Leu Pro Cys Val Arg Val Gly Asn Gln Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Val Ala Val Ser Tyr Ile Gly Ala Pro Leu Asp Ser Leu Arg Arg His Val Asp Leu Met Val Gly Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Gly Ala Phe Leu Val Gly Gln Met Phe Ser Phe Gln Pro Arg Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Ala Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Thr Leu Leu Leu Ala Gln Val Met Arg Ile Pro

Ser Thr Leu Val Asp Leu Leu Ala Gly Gly His Trp Gly Val Leu

Val Gly Leu Ala Tyr Phe Ser Met Gln Ala Asn Trp Ala Lys Val 175 170 Ile Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala 190 185

INFORMATION FOR SEQ ID NO:94: (2)

SEQUENCE CHARACTERISTICS: (i)

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown (D)

ORIGINAL SOURCE: (vi)

> (A) ORGANISM: homosapiens

INDIVIDUAL ISOLATE: Z7

SEQUENCE DESCRIPTION: SEQ ID NO:94: (xi)

Val Asn Tyr His Asn Ala Ser Gly Val Tyr His Ile Thr Asn Asp Cys Pro Asn Ser Ser Ile Met Tyr Glu Ala Glu His His Ile Leu 25 20 His Leu Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Gln Ser 35 40 Arg Cys Trp Val Ala Leu Thr Pro Thr Val Ala Ala Pro Tyr Ile 55 Gly Ala Pro Leu Glu Ser Ile Arg Arg His Val Asp Leu Met Val 75 65

Gly Ala Ala Thr Val Cys Ser Ala Leu Tyr Ile Gly Asp Leu Cys 85 90 80 Gly Gly Val Phe Leu Val Gly Gln Met Phe Ser Phe Gln Pro Arg

105 95 100 Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Ala Gly 115 120 110

His Val Thr Gly His Arg Met Ala Trp Asp Met Met Asn Trp 135 130 125 Ser Pro Thr Thr Leu Val Leu Ala Gln Val Met Arg Ile Pro

145 140 Ser Thr Leu Val Asp Leu Leu Thr Gly Gly His Trp Gly Ile Leu 155 160 Ile Gly Val Ala Tyr Phe Cys Met Gln Ala Asn Trp Ala Lys Val

170 175 Ile Leu Val Leu Phe Leu Tyr Ala Gly Val Asp Ala 190

185

INFORMATION FOR SEQ ID NO:95: (2)

> SEQUENCE CHARACTERISTICS: (i) (A) LENGTH: 192 amino acids

TYPE: amino acid (B)

(C) STRANDEDNESS: unknown

35

5

10

15

20

2.5

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: DK13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

Tyr Asn Tyr Arg Asn Ser Ser Gly Val Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Thr Asp Tyr His Ile Leu 3.0 20 His Leu Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Lys Ser 35 40 Thr Cys Trp Val Ser Leu Thr Pro Thr Val Ala Ala Gln His Leu 10 50 55 Asn Ala Pro Leu Glu Ser Leu Arg Arg His Val Asp Leu Met Val 70 65 Gly Gly Ala Thr Leu Cys Ser Ala Leu Tyr Ile Gly Asp Val Cys 90 80 85 Gly Gly Val Phe Leu Val Gly Gln Leu Phe Thr Phe Gln Pro Arg 105 95 100 15 Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Thr Gly 110 115 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 125 Ser Pro Thr Ala Thr Leu Val Leu Ala Gln Leu Met Arg Ile Pro 140 145 Gly Ala Met Val Asp Leu Leu Ala Gly Gly His Trp Gly Ile Leu 160 155 20 Val Gly Ile Ala Tyr Phe Ser Met Gln Ala Asn Trp Ala Lys Val 170 175 Ile Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala

185
(2) INFORMATION FOR SEQ ID NO:96:

25 (i) SEQUENCE CHARACTERISTICS:

A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

C) INDIVIDUAL ISOLATE: SA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Val Pro Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp

35

30

0

Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Ser Leu Ile Leu 30 20 His Ala Pro Gly Cys Val Pro Cys Val Arg Gln Asp Asn Val Ser 40 Arg Cys Trp Val Gln Ile Thr Pro Thr Leu Ser Ala Pro Thr Phe 50 55 Gly Ala Val Thr Ala Pro Leu Arg Arg Ala Val Asp Tyr Leu Ala 70 65 5 Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Ala Cys 85 90 80 Gly Ala Val Phe Leu Val Gly Gln Met Phe Thr Tyr Arg Pro Arg 95 105 Gln His Thr Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly 115 120 110 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp 135 125 10 Ser Pro Thr Thr Ala Leu Leu Met Ala Gln Met Leu Arg Ile Pro 145 140 Gln Val Val Ile Asp Ile Ile Ala Gly Gly His Trp Gly Val Leu 155 160 165 Phe Ala Ala Ala Tyr Phe Ala Ser Ala Ala Asn Trp Ala Lys Val 175 170 Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Gly 15 185 190

## (2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

C) INDIVIDUAL ISOLATE: SA4

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

35

30

Gly Ala Val Phe Leu Val Gly Gln Met Phe Thr Tyr Arg Pro Arg Gln His Thr Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Thr Ala Leu Leu Met Ala Gln Leu Leu Arg Ile Pro Gln Val Val Ile Asp Ile Ile Ala Gly Gly His Trp Gly Val Leu Phe Ala Ala Ala Tyr Phe Ala Ser Ala Ala Asn Trp Ala Lys Val Ile Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala 

(2) INFORMATION FOR SEQ ID NO:98:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: SAS
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:
- Val Pro Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Asn Leu Ile Leu His Ala Pro Gly Cys Val Pro Cys Val Lys Glu Gly Asn Val Ser Arg Cys Trp Val Gln Ile Thr Pro Thr Leu Ser Ala Pro Asn Leu Gly Ala Val Thr Ala Pro Leu Arg Arg Val Val Asp Tyr Leu Ala Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Ala Cys Gly Ala Val Phe Leu Val Gly Gln Met Phe Thr Tyr Arg Pro Arg Gln His Thr Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Asn Trp Ser Pro Thr Thr Ala Leu Val Met Ala Gln Val Leu Arg Ile Pro Gln Val Val Ile Asp Ile Ile Ala Gly Gly His Trp Gly Val Leu

```
Phe Ala Val Ala Tyr Phe Ala Ser Ala Ala Asn Trp Ala Lys Val
170 175 180
Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Gly
185 190
```

### (2) INFORMATION FOR SEQ ID NO:99:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: SA6

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

Val Pro Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp 5 10 15

Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Leu Ile Leu 20 25 30

His Ala Pro Gly Cys Val Pro Cys Val Arg Lys Asp Asn Val Ser 35 40

Arg Cys Trp Val His Ile Thr Pro Thr Leu Ser Ala Pro Ser Leu 50 60

Gly Ala Val Thr Ala Pro Leu Arg Arg Ala Val Asp Tyr Leu Ala
65 70 75
Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Val Cys

80 85 90 Gly Ala Leu Phe Leu Val Gly Gln Met Phe Thr Tyr Arg Pro Arg 95 100 105

Gln His Ala Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly 110 115 120 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Tyr

Ser Pro Ala Thr Ala Leu Val Met Ala Gln Met Leu Arg Ile Pro 145
Gln Val Val Ile Asp Ile Ile Ala Gly Gly His Trp Gly Val Leu

Gin val val lie Asp lie lie Ala Gly Gly His irp Gly val Leu 155 160 165 Phe Ala Ala Ala Tyr Phe Ala Ser Ala Ala Asn Trp Ala Lys Val 170 175

Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala 185 190

(2) INFORMATION FOR SEQ ID NO:100:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 192 amino acids
    - (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown

35

30

10

15

20

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: SA7

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Val Pro Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Asn Leu Ile Leu 30 20 His Ala Pro Gly Cys Val Pro Cys Val Arg Gln Asn Asn Val Ser 35 40 Arg Cys Trp Val Gln Ile Thr Pro Thr Leu Ser Ala Pro Asn Leu 10 50 55 Gly Ala Val Thr Ala Pro Leu Arg Arg Ala Val Asp Tyr Leu Ala 70 65 Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Ala Cys 80 85 90 Gly Ala Val Phe Leu Val Gly Gln Met Phe Ser Tyr Arg Pro Arg 100 105 95 15 Gln His Thr Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly 115 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Asn Trp 125 130 135 Ser Pro Thr Thr Ala Leu Val Met Ala Gln Leu Leu Arg Ile Pro 140 145 Gln Val Val Ile Asp Ile Ile Ala Gly Gly His Trp Gly Val Leu 155 160 20 Phe Ala Ala Ala Tyr Phe Ala Ser Ala Ala Asn Trp Ala Lys Val 170 175 Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala

185
(2) INFORMATION FOR SEQ ID NO:101:

(i) SEOUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

190

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: SA13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

Val Pro Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp 5 10 15

35

25

30

Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Leu Ile Leu 25 20 His Ala Pro Gly Cys Val Pro Cys Val Arg Gln Gly Asn Val Ser 40 Arg Cys Trp Val Gln Ile Thr Pro Thr Leu Ser Ala Pro Ser Leu 55 Gly Ala Val Thr Ala Pro Leu Arg Arg Ala Val Asp Tyr Leu Ala 5 65 70 Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Ala Cys 85 90 Gly Ala Val Phe Leu Val Gly Gln Met Phe Thr Tyr Ser Pro Arg 95 100 105 Arg His Asn Val Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly 110 115 120 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Asn Trp 10 125 130 135 Ser Pro Thr Thr Ala Leu Val Met Ala Gln Leu Leu Arg Ile Pro 140 145 150 Gln Val Val Ile Asp Ile Ile Ala Gly Ala His Trp Gly Val Leu 160 155 Phe Ala Ala Ala Tyr Tyr Ala Ser Ala Ala Asn Trp Ala Lys Val 170 175 15 Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala 185

### (2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: HK2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

Leu Thr Tyr Gln Asn Ser Ser Gln Leu Tyr His Leu Thr Asn Asp 15 Cys Pro Asn Ser Ser Ile Val Leu Glu Ala Asp Ala Met Ile Leu 30 20 25 His Leu Pro Gln Cys Leu Pro Cys Val Arg Val Asp Asp Arg Ser 35 40 Thr Cys Trp His Ala Val Thr Pro Thr Leu Ala Ile Pro Asn Ala 5.0 Ser Thr Pro Ala Thr Gln Phe Arg Arg His Val Asp Leu Leu Ala 75 65 70 Gln Ala Ala Val Val Cys Ser Ser Leu Tyr Ile Gln Asp Leu Cys 85 90 80

35

30

20

25

Gln Ser Leu Phe Leu Ala Gln Gln Leu Phe Thr Phe Gln Pro Arg 100 95 Arg His Trp Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Thr Gln 120 110 115 His Val Thr Gln His Arg Met Ala Trp Asp Met Met Asn Trp 135 125 130 Ser Pro Thr Thr Leu Val Leu Ser Ser Ile Leu Arg Val Pro 5 140 145 Glu Ile Cys Ala Ser Val Ile Phe Gln Gln His Trp Gln Ile Leu 160 155 Leu Ala Val Ala Tyr Phe Gln Met Ala Gln Asn Trp Leu Lys Val 180 170 Leu Ala Val Leu Phe Leu Phe Ala Gln Val Glu Ala 190 185

10

15

0

#### INFORMATION FOR SEQ ID NO: 103: (2)

- (i) SEQUENCE CHARACTERISTICS:
  - 573 base pairs (A) LENGTH:
    - TYPE: nucleic acid (C)
      - STRANDEDNESS: single
    - linear (D) TOPOLOGY:
- ORIGINAL SOURCE: (vi)
  - (A) ORGANISM: homosapiens
    - (C) INDIVIDUAL ISOLATE: DK7
- SEQUENCE DESCRIPTION: SEQ ID NO: 103: (xi)

20	ΔTG	AGC	ACG	AAT	ССТ	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
				CGT										78
		GGT										CCG	CGC	117
	AGG	GGC	CCT	AGA	TTG	GGT	GTG	CGC	GCG	CCG	AGG	AAG	ACT	156
	TCC	GAG	CGG	TCG	CAA	CCT	CGA	GGT	AGA	CGT	CAG	CCT	ATC	195
	CCC	AAG	GCA	CGT	CGG	CCC	GAG	GGC	AGG	ACC	TGG	GCT	CAG	234
	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	TGC	273
25	GGG	TGG	GCG	GGA	TGG	CTC	CTG	TCT	CCC	CGT	GGC	TCT	CGG	312
	CCT	AGC	TGG	GGC	CCC	ACA	GAC	CCC	CGG	CGC	AGG	TCG	CGC	351
	AAT	TTG	GGT	AAA	GTC	ATC	GAT	ACC	CTT	ACG	TGC	GGC	TTC	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATA	CCG	CTC	GTC	GGC	GCC	CCT	429
	CTT	GGA	GGC	GCT	GCC	AGG	GCC	CTG	GCG	CAT	GGC	GTC	CGG	468
	GTT	CTG	GAA	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG	AAC	CTT	507
	CCT	GGT	TGC	TCT	TTC	TCT	ATC	TTC	CTT	TTG	GCC	CTG	CTC	546
20	TCT	TGC	CTG	ACC	GTG	CCC	GCT	TCG	GCC					573
30														

#### (2) INFORMATION FOR SEO ID NO: 104:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 573 base pairs
    - TYPE: nucleic acid

•		
	<pre>(C) STRANDEDNESS: single (D) TOPOLOGY: linear</pre>	
	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US11	
5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:	
10		39 78 117 156 195 234 273 312 351 390 429 468
15	GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAC CTT CT GGT TGC TCT TTC TCT ATC TTC CTT CTG GCC CTG CTC TCT TGC CTG ACT GTG CCC GCT TCA GCC	507 546 573
	(2) INFORMATION FOR SEQ ID NO: 105:	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S14	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:	
30	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT GGC GGT CAA TTG GTT GGT GGA GTT TAC TTG TTG CCG CGC AGG GGC CCT AGA TTG GGT GTG CGC GGG AGG AGG AGG ATC TCC GAG GGT CGG CCC AGG GCG ACG ACG ACG ACG ACG ACG	39 78 117 156 195 234 273 312 351 390 429 468
35		

۰			
	GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG A CCT GGT TGC TCT TTC TCT ATC TTC CTC CTA GCC C TCT TGC CTG ACT GTG CCC GCT TCA GCC		507 546 573
	(2) INFORMATION FOR SEQ ID NO: 106:		
5	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear		
10	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SW1		
	(xi) SEQUENCE DESCRIPTION: SEQ ID	NO: 106:	
15	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC A AAC AAC CAT CGC CCA CAG GAC GTC AAG TTC C GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG C AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AT TCC GAG CGG TCG CAA CCT CGA GGT AGA CGT CAG CCC AAG GCG CGT CGG CCC CAG GGC AGG ACC TGG G CCC GGG TAT CCT TGG CCC CTC TAT GGC AAT GAG G CCT AGC TGG GGC CCT ACA GAC CCC CGT GGC CT CCT AGC TGG GGC CCT ACA GAC CCC CGG CGT AGG T AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC G CCT AGC TGG GGC CCT ACC GGC CGT CGT GGC CT CTT GGA GGC CCC CGC CGC CGT CGC CGC CCT GAC CTC ATG CGC CCC CGC CGT CGC CGC CCT GGC GGC CTC ATG CCC CCC CGC CGC CGC CGC CCT GGC GCC CTC ATG CCC CCC CCC CGC CCC CCT GGC CTC ATG CCC AGG CCC TCC CTC CTC CGC CCC CCT CGT CGC CGC CGC CTC ACC CCC CCC CCC CCC CCC CCT CGT TGC TCT TCC TCT ACC TCC CTC CTC CCC CCC CCT CGT TGC TCT TCC TCT ATC TCC TCT CTC CCC CCC CCT CCT CCC CCC C	CG GGT CG CGC ACT CT ATC CT CAG GC TGC CT CGG CG CGC GC TTC CC CCT TC CGG ACC ACC ACC ACC ACC ACC ACC ACC ACC	35 71 15 15 19 23 31 33 35 31 46 50 51 57
	(2) INFORMATION FOR SEQ ID NO: 107:		57.
25	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear		
30	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S18		
	(xi) SEQUENCE DESCRIPTION: SEQ ID	NO: 107:	
35	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAAC ACC AAC CGT CGC CCA CAG GAC GTT AAG TTC C		3 ! 7 !

372577\_1

```
GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC
     AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT
     TCC GAG CGG TCG CAA CCT CGC GGT AGA CGT CAG CCT ATC
                                                                195
     CCC AAG GCG CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG
                                                                234
     CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TGC
                                                                273
     GGG TGG GCG GGA TGG CTC CTG TCC CCC CGT GGC TCC CGG
                                                                312
     CCT AGC TGG GGC CCT ACA GAC CCC CGG CGT AGG TCG CGC
                                                               351
     AAT TTG GGC AAA GTC ATC GAT ACC CTC ACG TGC GGC TTC
                                                               390
     GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCT
                                                               429
     CTC GGA GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG
                                                               468
     GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAC CTT
                                                               507
     CCT GGT TGC TCT TTC TCT ATC TTC CTT CTG GCC CTG CTC
                                                               546
     TCT TGT CTG ACT GTG CCC GCG TCA GCT
                                                               573
10
     (2)
           INFORMATION FOR SEO ID NO: 108:
                      SEQUENCE CHARACTERISTICS:
             (i)
                      (A)
                           LENGTH: 573 base pairs
                      (B)
                           TYPE: nucleic acid
                      (C)
                           STRANDEDNESS: single
                      (D)
                           TOPOLOGY:
                                      linear
15
                      ORIGINAL SOURCE:
             (vi)
                      (A)
                           ORGANISM: homosapiens
                      (C)
                           INDIVIDUAL ISOLATE:
             (xi)
                      SEQUENCE DESCRIPTION: SEQ ID NO: 108:
                                                                 39
     ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT
20
     AAC ACC AAC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT
                                                                78
     GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC
                                                                117
     AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT
                                                                156
     TCC GAG CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT ATC
                                                                195
     CCC AAG GCG CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG
                                                                234
     CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TGC
                                                                273
     GGG TGG GCG GGA TGG CTC CTG TCC CCC CGT GGC TCT CGG
                                                                312
25
     CCT AGC TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGC
                                                                351
     AAT TTG GGT AAG GTC ATC GAC ACC CTC ACG TGC GGC TTC
                                                               390
     GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCC CCC
                                                               429
     CTT GGG GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGA
                                                               468
     GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAT CTT
                                                               507
     CCT GGT TGC TCT TTC TCT ATC TTC CTT TTG GCT TTG CTC
                                                                546
                                                                573
     TCT TGC TTG ACC GTG CCC GCA TCG GCC
30
      (2)
            INFORMATION FOR SEO ID NO: 109:
                      SEQUENCE CHARACTERISTICS:
             (i)
                      (A) LENGTH: 573 base pairs
```

TYPE: nucleic acid

STRANDEDNESS: single

(B)

372577 1

٥			
		(D) TOPOLOGY: linear	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA10	
5	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 109:	
10	AAC ACC AAC CGC GGT GGT CAG ATC AGG GGC CCC AGG TCC GAG GCT CGC CCC AAG GCT CGC CCC GGG TAC CCT GGG TGG GGA GGA CCT AGT TGG GGC AAT TTG GGT AAG	CCT AAA CCT CAA AGA AAA ACC AAA CGT CGC CCA CAG GAC GTC AAG TTC CCG GGC GTT GGT GGA GTC TAT CTG TTG CCG CGC TTG GGT GT CGC GCG ACG AGG AAG ACT CAA CCT CGT GGA AGG CGA CAG ACC TATC CAG CCC GAG GGC AGG ACC TGG GCC CAG TGG CCC CTC TAT GGC AAT GAG GGC TTG TGG CTC CTG TCA CCC CGT GGT TCG CGT CCC AGG GAC CCC CGG CGT AGG TCG CGT GTC ATC GAT ACC CTC ACA TGC GCC TCC CTC ATC GAT ACC CTC ACA TGC GCC CTC	39 78 117 156 195 234 273 312 351 390
15	TTA GGG GGC GCT GTT CTG GAA GAC CCC GGT TGC CCT		429 468 507 546 573
	(2) INFORMATI	ON FOR SEQ ID NO: 110:	
20	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S45	
25	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 110:	
30	AAC ACC AAC GGG GGC GGT CAG ATC AGG GGC CCC AGG TCC GGG GGT CAC CCC AGG CGT CCC GGG TGG GCA GGA CCT AGT TGG GGC AAT TTG GGC AAC GCC GAC CTC ATG CTA AGG GGC GGC TCT AGG GGC GGC TCT AGG GGC GGC AGG GGC GGC AGG GGC GGC	CCT AAA CCT CAA AGA CAA ACC AAA CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT GTT GGT GGA GTT TAC CTG TTG CCG CGC TTG GGT GTG CGC GCG ACT AGG AAA ACT CAA CCT CGT GGA CGG CGA CAA CCT ATC CGG CCC GAG GGC AGG GCC TGG GCC CAG TGG CCC CTC TAT GGC AAT GAG GGC TTG TGG CTC CTC TAT GGC AAT GAG GCC TTG CCC ACG GAC CCC CGG CGT AGG TCC CGG CCC ACG GAC CCC CGG CGT AGG TCC CGC GTC ATC GAT ACC CTC AGG TGC GCC TTC GGC AGA GCC TTG GCC CGC CCC GGC AGA GCC TTG GCC CCC GGC AGA GCC TTG GCC CTC GGC CTC GGC GTG AAC TAT GCA ACA GGG AAT CTG	39 78 117 156 234 273 312 351 390 429 468 507
35	372577_1		

0		
	CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT CTG CTG TCC TGC TTG ACC ATC CCA GCT TCC GCT	546 573
	(2) INFORMATION FOR SEQ ID NO: 111:	
5	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
10	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: D1	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:	
15 20	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC GGT GGT CAG TCG GTT GGT GGA GTT TAC CTG TTG CGC GGC AGG GGC CCC AGG GTC CAA CCT CGT TGA AGG CAC CTG ACT CCC AAG GCT CGC AAC CTC CTT GGA AGG CGA CAA CCT ATC CCC AAG GGT CGC CAA CCT CGT GGA AGG CGA CAA CCT ATC CCC AAG GGT CGC CAG CCC GAG GGT AGG GCC TGG GCT CAG CCC GGG TAC CCT TGC CCC ACC CGC GGC TCC CTG TCC CAG CCC ACC GGC TCC CGG CCT CGT TCA CCC CGC GGC TCC CGG CCT AGT TGG GCC CAC GAC CCC CGG CGC TCC CGG CCC CGG CTC CGC CCT AGT TGG GGC TAC ATC CCC CGC GGC TCC CGG CCT AGT TGG GGC CCC ACC GAC CCC CGC CTC ACT CGC CCC CGC CCC CGG CCT CGT CGC CCC CGC CCC CTC GTC CGC CCC CC	39 78 117 156 195 234 273 312 350 429 468 507 546 573
25	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
30	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US6	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:	
35	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT	39 78 117 156

372577\_1

0															
5	CCC CCC GGG CCT AAT GCC CTA GTT CCC	AAG GGG TGG AGT TTG GAC GGG CTG GGT	GCT TAC GCA TGG GGT CTC GGC GAG	CGC CCT GGA GGC AAG ATG GCT GAC TCT	CGG TGG TGG CCC GTC GGG GCC GGC TTC	CCC CCC CTC ACG ATC TAC AGG GTG TCT	GAG CTC CTG GAC GAT ATT GCC AAC ATC	GGC TAT TCA CCC ACC CCG TTG TAT TTC	AGG GGC CCC CGG CTC CTC GCG GCA CTC	GCC AAC CGT CGT ACA GTC CAT ACA	CAA TGG GAG GGC AGG TGC GGC GGC GGC GGC	GCT GGC TCG TCG GGC GCC GTC AAC	CAG ATG CGG CGT TTC		195 234 273 312 351 390 429 468 507 546 573
10	(2)		NFORI	ITAN	SEQUAL (A)	JENCI LEI TYI STI	E CHA NGTH PE: RANDI	ARACT 5 5 1 nucl	TERIS 73 ba leic SS:	STICS ase p acid	pairs d	5			
15			(vi)		(A) (C)	GINAI OR( IN)		JRCE SM: DUAL	homo	osap: LATE	: P:		D: 113	3:	
20	AAC GGT AGG TCC CCC CCC	ACC GGT GGC GAG AAG GGG	AAC CAG CCC CGG GCT TAC	CGC ATC AGG TCG CGC CCT	CGC GTT TTG CAA CGG TGG	CCA GGT GGT CCT CCC	CAG GGA GTG CGT GAG CTC	GAC GTT CGC GGA GGC TAT	GTC TAC GCG AGG AGG GGC	AAG CTG ACT CGA GCC AAT	ACC TTC TTG AGG CAA TGG GAG GGC	CCG CCG AAG CCT GCT GGC	GGC CGC ACT ATC CAG TTG		39 78 117 156 195 234 273 312
25	CCT AAT GCC CTA GTT CCC	AGT TTG GAC GGG CTG GGT	TGG GGT CTC GGC GAG	GGC AAG ATG GCT GAC TCT	CCC GTC GGG GCC GGC TTC	ACG ATC TAC AGG GTG TCT	GAC GAT ATT GCC AAC ATC	CCC ACC CCG CTG TAT TTC	CGG CTC CTC GCG GCA CTC	CGT ACA GTC CAT ACA	AGG TGC GGC GGC GGG	TCG GGC GCC GTC AAT	CGT TTC CCC CGG CTG		351 390 429 468 507
	(-)	_				an a									

#### INFORMATION FOR SEQ ID NO: 114: (2) 30

- (i)
- SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 573 base pairs
  (B) TYPE: nucleic acid

  - (C) (D) STRANDEDNESS: single TOPOLOGY: linear

```
(vi)
                     ORIGINAL SOURCE:
                     (A)
                          ORGANISM: homosapiens
                     (C)
                          INDIVIDUAL ISOLATE: DK1
            (xi)
                     SEQUENCE DESCRIPTION: SEQ ID NO: 114:
     ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT
                                                                39
     AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC
                                                                78
                                                               117
     GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC
                                                               156
     AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT
     TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC
     CCC AAG GCT CGC CGG CCC GAG GGC AGG GCC TGG GCT CAG
                                                               234
     CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC ATG
                                                               273
     GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGC TCT CGG
                                                               312
     CCT AGT TGG GGC CCC AAC GAC CCC CGG CGT AGG TCG CGT
                                                               351
10
     AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC
                                                               390
     GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC
                                                               429
     CTA GGG GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG
     GTT CTG GAG GAC GGC GTG AAC TAC GCA ACA GGG AAT TTG
                                                               507
     CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT CTG TTG
                                                               546
     TCC TGT TTG ACC ATC CCA GCT TCC GCC
                                                               573
15
           INFORMATION FOR SEQ ID NO: 115:
     (2)
                     SEQUENCE CHARACTERISTICS:
            (i)
                         LENGTH: 573 base pairs
                      (A)
                      (B)
                          TYPE: nucleic acid
                      (C)
                          STRANDEDNESS: single
                      (D)
                          TOPOLOGY:
                                     linear
20
            (vi)
                     ORIGINAL SOURCE:
                      (A)
                          ORGANISM: homosapiens
                      (C)
                           INDIVIDUAL ISOLATE: T10
                     SEQUENCE DESCRIPTION: SEQ ID NO: 115:
            (xi)
     ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT
                                                                39
     AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC
2.5
                                                                78
     GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC
                                                               117
     AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT
                                                               156
     TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAG CCT ATC
     CCC AAG GCT CGC CAG CCC GAG GGC AGG GCC TGG GCT CAG
     CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC ATG
                                                               273
     GGG TGG GCA GGA TGG CTC CTG TCA CCC CGT GGC TCC CGG
                                                               312
     CCT AGT TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGT
                                                               351
30
     AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC
                                                               390
     GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC
                                                               429
     CTA GGG GGC GCT GCC AGG GCT CTG GCA CAT GGT GTC CGG
                                                               468
     GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT TTG
                                                               507
     CCC GGT TGC TCT TTT TCT ATC TTC CTC TTG GCT CTG CTG
                                                              546
     TCT TGT CTG ACC ATC CCA GCT TCC GCT
                                                               573
```

	(2) INFORMATION FOR SE	EQ ID NO: 116:	
5	(A) LEI (B) TYI (C) STE	G CHARACTERISTICS: NGTH: 573 base pairs PE: nucleic acid RANDEDNESS: single POLOGY: linear	
	(A) ORG	L SOURCE: GANISM: homosapiens DIVIDUAL ISOLATE: SW2	
	(xi) SEQUENCE	E DESCRIPTION: SEQ ID NO: 116:	
10	AAC ACC AAC CGC CGC CCA GGT GGC CAG ATC GGT GGT AGG GGC CCC CGG TTG GGT TCC GAG CGG TCG CAA CCT CCC AAG GCT CGC CAG CCC CCT GGG TAC CCC TGG CCC	CCT CAA AGA AAA ACC AAA CGT CAG GAC GTC AAG TTC CCG GGC GGA GTT TAC CTG TTG CCG CGC GTG CGC GCG ACT AGG AAG ACT CGT GGA AGG CGA CAA CCT ATC GAG GGC AGG GCC TGG GCT CAG CTC TAT GGC AAT GAG GGC ATG	39 78 117 156 195 234 273
15	GGA TGG GCA GGA TGG CTC CCT AGT TGG GGC CCC ACT AAT TTG GGT AAG GTC ATC GCC GAC CTC ATG GGG TAC CTA GGG GGC GCT GCC AGG GTC CTG GAG GAC GGC GTG	CTG TCC CCC CGC GGC TCT CGG GAC CCC CGG CGT AGG TCG GGT GAT ACC CTC ACA TGC GGC TTC ATT CCG CTC GTC GGC GCC CCC GCC CTG GCG CAT GGC GTC CGG AAC TAT GCA ACA GGG AAT CTG ATC TTC CTC TTG GCT TTG CTG	312 351 390 429 468 507 546 573
20	100 101 010 1100 1110 1111		
	(2) INFORMATION FOR S	EQ ID NO: 117:	
25	(A) LE (B) TY (C) ST	E CHARACTERISTICS: NGTH: 573 base pairs PE: nucleic acid RANDEDNESS: single POLOGY: linear	
	(A) OR	L SOURCE: GANISM: homosapiens DIVIDUAL ISOLATE: IND3	
30	(xi) SEQUENC	E DESCRIPTION: SEQ ID NO: 117:	
35	AAC ACC AAC CGC CGC CCA GGT GGC CAG ATC GTT GGT AGG GGC CCC AGG TTG GGT TCC GAG CGG TCG CAA CCT	CCT CAA AGA AAA ACC AAA CGT CAG GAC GTC AAG TTC CCG GGC GGA GTT TAC CTG TTG CCG CGC GTG CGC GCG ACT AGG AAG ACT CGT GGA AGG CGA CAA CCT ATC GAG GGT AGG GCC TGG GCT CAG	39 78 117 156 195 234

CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TTG 273 GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGT TCT CGG 312 CCT AGT TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGT 351 AAT TTG GGT AAA GTC ATC GAT ACC CTC ACA TGC GGC TTC 390 GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCC CCC 429 CTA GGG GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG 468 GTC CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAC TTG 507 5 CCC GGT TGC TCT TTC TCT ATC TTC CTT TTA GCT TTG CTA 546 TCC TGT TTG ACC ATC CCA GCT TCC GCT 573 INFORMATION FOR SEQ ID NO: 118: (2) (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs TYPE: nucleic acid (B) 10 (C) STRANDEDNESS: single (D) TOPOLOGY: linear (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: 15 SEQUENCE DESCRIPTION: SEQ ID NO: 118: (xi) ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT 39 AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC 78 GGT GGC CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT 156 TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC 195 CCC AAG GCT CGC CGG CCC GAG GGT AGG GCC TGG GCT CAG 234 20 CCC GGG CAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TTG 273 GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGC TCT CGG 312 CCT AGT TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGT 351 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC 390 GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCC CCC 429 CTA GGG GGT GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG 468 GTC CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAC TTG 507 CCC GGT TGC TCT TTC TCT ATC TTC CTT TTG GCT TTG CTA 546 25 TCC TGT TTG ACC GTC CCA GCT TCC GCT 573 (2) INFORMATION FOR SEQ ID NO: 119: SEQUENCE CHARACTERISTICS: (i) (A) LENGTH: 573 base pairs 30 TYPE: nucleic acid (B) (C) STRANDEDNESS: single TOPOLOGY: linear (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens INDIVIDUAL ISOLATE: S9 (C) 35

SEQUENCE DESCRIPTION: SEO ID NO: 119: (xi) ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT 39 AAC ACC AAC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGC 78 GGT GGT CAG ATC GTC GGT GGA GTT TAC CTG TTG CCG CGC 117 AGG GGC CCC AGG TTG GGT GTG CGC GCA ACT AGG AAG ACT 156 5 TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC 195 CCC AAG GCT CGC CAT CCC GAG GGC AGG GCC TGG GCT CAG 234 CCC GGG TAC CCT TGG CCC CTC TAC GGC AAT GAG GGC TTG 273 GGG TGG GCA GGA TGG CTC CTG TCA CCC CGT GGC TCT CGG CCT AGT TGG GGC CCC AAT GAC CCC CGG CGT AGG TCG CGT 351 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTT 390 GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC 429 CTA GGG GGC GCT GCC AGG GCT CTG GCG CAT GGC GTC CGG 468 10 GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAC CTC 507 CCC GGT TGC TCT TTC TCT ATC TTC CTT CTG GCT TTG CTG 546 TCC TGT TTG ACC ATC CCA GCT TCC GCT 573 INFORMATION FOR SEQ ID NO: 120: (2) 15 SEQUENCE CHARACTERISTICS: (i)(A) LENGTH: 573 base pairs TYPE: nucleic acid (B) (C) STRANDEDNESS: single (D) TOPOLOGY: linear ORIGINAL SOURCE: (vi) (A) ORGANISM: homosapiens 20 (C) INDIVIDUAL ISOLATE: HK3 SEQUENCE DESCRIPTION: SEQ ID NO: 120: (xi) ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT 39 78 AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC 117 GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC AGG GGC CCC AGG TTG GGT GTG CGC GCG ACC AGG AAG ACT 156 25 TCA GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC 195 CCC AAG GCT CGC CAA CCC GAG GGC AGG ACC TGG GCT CAG 234 CCC GGG TAT CCT TGG CCC CTC TAT GGC AAC GAG GGC ATG 273 GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGC TCT CGG CCT AAT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGC 351 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC 390 GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGT GCC CCC 429 30 468 CTA GGG GGC GTT GCC AGA GCC TTG GCA CAT GGT GTC CGG GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT TTA 507 CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG CTG 546 573 TCC TGC TTG ACC ACC CCA GCT TCC GCT

(2) INFORMATION FOR SEQ ID NO: 121:

5	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
J	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK5	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 121:	
10	AAC ACC AAC CGC GGT GGT CAG ATC AGG GGC CCC AGG TCC GAG CGG TCG CCC AAG GCT CGC	CCT AAA CCT CAA AGA AAA ACC AAA CGT CGC CCA CAG GAC GTC AAG TTC CCG GGC GTT GGT GGA GTT TAC CTG TTG CGC GCT TTG GGT GTG GGC ACC AGG AAG ACT CAA CCT CGT GGA AGG CGA CAA CCT ATC CGA CCC GAG GGC AGG ACC TGG GCT CAG	39 78 117 156 195 234
15	GGG TGG GCA GGA CCT AGT TGG GGC AAT TTG GGT AAG GCC GAC CTC ATG CTA GGG GGC GTT GTT CTG GAG GAC CCC GGT TGC TCT	TGG CCC CTC TAT GGC AAT GAG GGC ATG TGG CTC CTG TCA CCC CAT GGC TCT CGG CCC ACG GAC CCC CGG CGT AGG TCG CGT GTC ATC GAT ACC CTC ACG TGC GGC TTC GGG TAC ATC CCG CTC GTC GGC GCC CCC GGC AGA GCC CTG GCA ACA GGG GAT CTC GGC GTG AAC TAC GCA ACA GGG AAT ATA TTC TCT ATC TTC CTT TTG GCT TTG CTG ACC CCA GTT TCC GCT	273 312 351 390 429 468 507 546 573
20	(2) INFORMATI	ON FOR SEQ ID NO: 122:	
25	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
25	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK4	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 122:	
30	AAC ACC AAC CGC GGT GGC CAG ATC AGG GGC CCC AGG TCC GAG GCT CGC CCC AAG GCT CGC CCC GGG TAC CCT	CCT AAA CCT CAA AGA AAG ACC AAA CGT CGC CCA CAG GAC GTT AAG TTC CCG GGC GTC GGT GGA GTT TAC CTG TTG CCG CGC TTG GGT GTG CGC GCG ACT AGG AAG ACT CAA CCT CGT GGA AGG CGA CAA CCT ATC CAA CCC GAG GGC AGG ACC TGG GCT CAG TGG CCC CTC TAT GGC AAT GAG GGC ATG ATG CTC CTG TGA CCC CGC GGC TCT CGG	39 78 117 156 195 234 273 312
35			

372577\_1

۰		
5	CCT AGT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGC AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGG GGC TTC GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC TTA GGG GGC GTT GCC AGA GCC CTG GCA CAT GGT GTC CGG GTT GTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT TTG CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT CTG TCC TGT TTG ACC ATC CCA GCT TCC GCT	351 390 429 468 507 546 573
	(2) INFORMATION FOR SEQ ID NO: 123:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: P8  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:	
15	ATG AGC ACG ACT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AGC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGC GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CGC GGC AGG GGC CCC AGG TTG GGT GGC GCG ACT AGG AAG ACT TCC GAG CGA TCG CAA CCT CGT GGC AGG CGA CAA CCT ATC CCC AAG GCT CGC CGG CCC GAG GGT AGG GCC TGG GCC CAG	39 78 117 156 195 234 273
20	CCC GGG CAC CCT TGG CCC CTC TAT GCC AAT GAG GGC TTG GGG TGG GGC GGC CTC CTG TCA CCC CGC GGC TCC CGG CCT AGT TGG GGC CCC ACG GAC CCC CGC GGT AGG TCG CGC AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GGC CCC CTA GGG GGC GTT GCC AGG GCC CTG GGC CAT GGC GTC CGG GTT GTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT CTG CCT GGT TGC TCT TTC TCT ATC TTC CTT TTG GCT TTG CTG TCT TGT CTG ACC ATC CCA GCT TCC GCT	312 351 390 429 468 507 546 573

30

25

	(2) INFORMATIO	ON FOR SEQ ID NO: 124:	
5	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: T3	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 124:	
10	AAC ACC AAC CGC GGT GGT CAG ATC AGG GGC CCC AGG TCC GAG CGG TCG	CCT AAA CCT CAA AGA AAA ACC AAA CGT CGC CCA CAG GAC GTT AAG TTC CCG GGC GTT GGT GGA GTT TAC CTG TTG CGC CGC TTG GGT GTG CGC GCG ACT AGG AAG ACT CAA CCT CGT GGA AGG CGA CAA CCT ATC CGG CCC GAG GGT AGG GCC TGG GCT CAG	39 78 117 156 195 234
15	CCC GGG TAC CCT GGG TGG GCA GGA CCT AAT TGG GGC AAT CTG GGT AAG GCC GAC CTC ATG TTA GGG GGC GTT GTT CTG GAG GAC	TGG CTC CTC TAT GGC GAC GAG GGC ATG TGG CTC CTG TCA CCC CGC GGC TCC CGG CCC ACA GAC CCC CGG GGT AGG TCC CGT GTC ATC GAT ACC CTC ACA TGC GGC TTC GGG TAC ATT CCG CTC GTC GGC GTC CCC GCC AGG GCC CTG GCG CAT GGC GTC CGG GGC GTG AAT TAC GCA ACA GGC AT TTG TTC TCT ATC CTC CTC TTG GCT TTG CTG	273 312 351 390 429 468 507 546
20	TCC TGC TTG ACC	ATC CCA GCT TCC GCT	573
	(2) INFORMATI	ON FOR SEQ ID NO: 125:	
25	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: T4	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 125:	
30	AAC ACC AAC CGT GGC GGC CAG ATC AGG GGC CCC AGG TCG GAG CGA TCC CCC AAA GAT CGG	CCT AAA CCT CAA AGA AAA ACC AAA AGA CGC CCA CAG GAC GTT AAG TTC CCG GGC GTT GGC GGA GTA TAC TTG TTG CCG CGC TTG GGT GTG CGC GCG ACA AGG AAG ACT CAG CCA CGT GGG AGG CGC CAG CCC ATC CGC TCC ACT GGC AAG TCC TGG GGA AAA	39 78 117 156 195 234
35	CCA GGA TAT CCC	TGG CCC CTG TAT GGG AAT GAG GGA CTC	273

CCC   TCC   TGG   GGC   CAAT   GAC   CCC   CGG   CAT   AGG   TGG   CGC   AAC   GTG   GTG   AAG   GTC   ATC   GAT   ACC   CTA   ACG   TGC   AGC   CTT   GGT   GGC   GAC   CTC   ATG   GGG   TAC   GTC   CCC   GTC   GTA   GGC   GGC   CCG   GTT   GGT   GGC   GTC   GGG   GTC   GTC   GGG   GTC   GTC   GGG   GTC   GTC   GGG   GTC	0															
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US10  (Xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:  ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACT AAC CGT CGC CCA CAA GAC GTT AAG TTT CCG GGC 7 AGG GGC CCA CAG GTT GGC GCA CAA GAC ACA AGA AAA ACC AAA AGA AAC ACT AAC CGT TGGC CCA CAA GAC GTT AAG TTT CCG GGC 7 AGG GGC CCA CAG CCC CCA CAA GGC ACA AGG AAG ACT 15 TCG GAG CGG TCC CAG CCA CCT GGG AGG CCC ATC 19 CCC AAA GAT CGT GGC CCC ACT GGC AAC AGG AAG ACT 15 CCA GGA TAC CCT TGC CCC CTA TAT GGG AAT GAC GGA CCC ATC 19 CCC AGG TCC CAG CCA CCT GGC AAC TCC CGG GAC ACA AGG AAAA ACC AAA AGA CCC CCC	5	CCC AAC GCC TTG GTC CCT	TCC GTG GAC GGT CTG GGT	TGG GGT CTC GGC GAG TGC	GGC AAG ATG GTC GAC TCC	CCC GTC GGG GCC GGG TTT	AAT ATC TAC AGA GTT TCT	GAC GAT GTC GCT AAT ATT	CCC ACC CCC CTC TAT TTC	CGG CTA GTC GCG GCA TTG	CAT ACG GTA CAT ACA	AGG TGC GGC GGC GGG	TCG AGC GGC GTG AAC	CGC CTT CCG AGA TTA		312 351 390 429 468 507 546 573
(A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US10  (Xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:  ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACT AAC CGT CGC CCA CAA GAC GTT AAG TTT CCG GGC 7 GGC GGC CAG ATC GGT GGC GCA CAA GAC AAA ACC AAA AGA AAC ACT AAC CGT TGGC CCA CAA GAC GTT AAG TTT TCG GGC 7 GGC GGC CAG ATC GGT GGT GGC GC CCA CAA AGA AAA ACC AAA ACT 15 TCG GAG CGG TCC CAG CCA CGT GGC ACA AGA AAC 115 TCG GAG CGG TCC CAG CCC CTA TAT GGG AAC CCC ATC 19 CCC AAA GAT CCC TGG CCC CCTA TAT GGG AAT CCC TGG GGA AAA ACC 125 CCC AGG TTG GGC CCC CTA TAT GGG AAT CCC TGG GCA CCC ATC 19 CCC AGG TGG GCC CCC CTA TAT GGG AAT CCC CGT 35 CCC TCT TGG GGC CCC CAT GGC CCC CGA GGT TCC CGT 35 AAC GTG GGT AAG GTC ATC GGT CCC CG CAT GGG TTG GGC CTC TGG GCC CCC ATC GGT ACC CCC GTC TGG GCC CTC TGG GCC CCC ATC GGT ACC CCC GTA TGG TGC GCC GTG AGC CTC ATC GGG AAC ACC CTC TGG GCC CCC ATC GGT ACC CCC GTA TGC GGC CTC TGG GCC GCC		(2)	II	NFORM	MATIO	ON FO	OR SI	EQ II	ON C	: 12	5:					
(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US10  15  (Xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:  ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACT AAC CGT CGC CCA CAA GAC AGT AAG TTT CCG GGC 7 GGC GGC CAG ATC GTT GGC GGA GTT AAG TTT CCG GGC 7 TCG GAG CGG TCC CAC CAA GAC ATT TTT TTG CCG CGC 11 AGG GGC CCC AGG TTG GGT GTG CGC CG ACA AGG AAG ACT 12 CCC CAAA GAT CGT CCC CCG CAT CGG AGC CCC ATC 19 CCC GAG ATC CCT TGG CCC CTA TAT GGA AT GA GGA CTC 27 GGC TGG GGA CGC CCC ACT GAT CCC CGC CAT CCC CGT 31 CCC TCT TGG GGC CCC ACT GAT CCC CGC CAT AGG TCC CCT 31 AAC GTG GGT AAG GTC ATC GAT ACC CTA AGG TCC CGC CTT 32 ACC GGC TCC CGC CAT GAT CCC CGC CAT AGG TCC CGC CTT GGT GGC CCC TAT CCC CGC CAT GGT AGG TCC CCC CGG CAT GGC CGC CAT GGC CCC CGG CAT CCC CGC CAT GGC CCC CGG CAT CCC CGC CAT GGC CCC CGG CAT CCC CGC CAT GGC CCC CGG CAT GGC CCC CCC CGG CAT GGC CCC CCG CAT GGC CCC CCC CGG CAT GGC CCC CCC CGG CAT GGC CCC CGG CAT GGC CCC CCCCC CGG CAT GGC CCC CCC CGC CAT CCC CCC CGC CAT GGC CCC CCC CGC CAT GGC CCC CCC CGC CAT GGC CCC C	10			(i)		(A) (B) (C)	LEI TYI STI	NGTH PE: RAND	nuci nuci	73 ba leic SS:	ase p acio sino	pair: d	3			
ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACT AAC CGT CGC CCA CAA GAC GTT AAG TTT CCG GGC GGC GGC CAG ATC GGT GGC GCA CAA GAC GTT AAG TTT CCG GGC TCG GGC GGC CAG GTT GGT GGC GCG CAC AGG AGG GGC CCC AGG TTG GGT GGC GCG CAC ACG CCC AAA GAT CCT TGG CCC CCA CT GGC ACA AGG AAG ACT CCC AGA TAC CCT TGG CCC CCT ATAT GGG AAT CCC GGC ATC CCC AGG TAC CCT TGG CCC CTA TAT GGG AAT CCC GGC ACA CCC TCT TGG GGC CCC CTA TAT GGG AAT CCC CGC CAT GCC CCC TCT TGG GGC CCC CACT GGC AAT CCC CGC CAT GGG CAC ACG ACG TGC GCC CCC ACT GGC AAT CCC CGC CAT GCG CAC ACG CCC TCT TGG GGC CCC ACT GGT CCC CGC CAT GGC TCC CGT AAC GTG GGT AAG GTC ATC GAT ACC CTA ACG TGC CGC ACC TCT TGG GGC CCC ACT GGT CCC CGC ATA GGT TCC CGC AAC GTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GCC TTT AGG GCC GAC TCC TGC GCC CAT GGC CTC CGC CTT TGG TGG CCC CAGA GCT CCC CGT CTG GGC CTC CGC CTT TGGT GGC CCC CAGA GCT CCC CGTC TGT GGC CTC CGC CTT TGGT GGC CTC GCC AGA GCT CTC GCC CAT GGC CTC CGC CTT TGGT GCC TCC GCC AGA GCT CTC GCC CAT GGC CTC CGC CTC TGG GCC CTC TTC TCT ACT CTC TCC CCC CTC TTC T	15					(A) (C)	ORG	GANI: DIVI	SM: DUAL	home ISO	LATE	: U		2. 126		
AAC ACT AAC CGT CGC CCA CAA GAC GTT AAG TTT CCG GGC GGC GGC CGG CAC ACC GTT GGC GGC ACG ACG CGC GGC GAC ACG ACG ACG	13					_									) :	
		AAC GGC AGG CCA GGC CCA GGC CCC AAC GCC CTC CCC CCC CCC CCC CCC CCC CCC CC	ACT GGC GGC GAG AAA GGA TGG TCT GTG GAC GGT CTG	AAC CAG CCC CGG GAT TAC GCA TGG GGT CTC GGC GAG	CGT ATC AGG TCC CGG CCT GGA GGC AAG ATG GTC GAC	CGC GTT TTG CAG CGC TGG TGG CCC GTC GGA GCC GGG	CCA GGC GGT CCA CCC CTC ACT ATC TAC AGA GTT TCT	CAA GGA GTG CGT ACT CTA CTG GAT ATC GCT AAT	GAC GTA CGC GGG GGC TAT TCC CCC CCC CTC TAT TTC	GTT TAC GCG AGG AAG GGG CCC CGG CTA GTC GCG GCA	AAG TTG ACA CGC TCC AAT CGA CAT ACG GTG CAT ACA CTG	TTT TTG AGG CAG TGG GAG GGT AGG TGC GGC GGC	CCG CCG AAG CCC GGA TCC TCG GGC GCT GTG AAC	GGC CGC ACT ATC AAA CTC CGT CGC TTT CCG AGG TTA		78 117 156 195 234 273 312 351 390 429 468 507 546
		(0)	_			ov. 5	op a	TO T	D NO	10	2					

### (2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: T9

(C) INDIVIDO

30

SEQUENCE DESCRIPTION: SEQ ID NO: 127: (xi) ATG AGC ACA AAT CCA AAA CCC CAA AGA AAA ACC ATA AGA 39 AAC ACC AAC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGC 78 GGC GGC CAG ATC GTT GGC GGA GTA TAC TTG TTG CCG CGC 117 AGG GGC CCT AGG TTG GGT GTG CGC ACG ACA AGG AAG ACT 156 5 TCG GAG CGG TCC CAG CCA CGT GGG AGG CGC CAG CCC ATC 195 CCC AAA GAT CGG CGC TCC ACT GGC AAG TCC TGG GGA AAA 234 CCA GGA TAC CCC TGG CCT CTA TAT GGG AAT GAG GGA CTC 273 GGC TGG GCG GGA TGG CTC CTG TCC CCC CGA GGT TCC CGT 312 CCC TCT TGG GGC CCC AGT GAC CCC CGG CAT AGG TCG CGC 351 AAC GTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGC TTT 390 GCC GAC CTC ATG GGG TAC ATC CCC GTC GTA GGC GCC CCG 429 CTT GGT GGC GTT GCC AGA GCT CTC GCG CAC GGC GTG AGA 468 10 507 GTC CTG GAG GAC GGG GTT AAT TAT GCA ACA GGG AAC CTA CCT GGT TGC TCT TTT TCT ATC TTC TTG CTG GCC CTA CTG 546 573 TCC TGC ATC ACC ACT CCG GCC TCT GCT INFORMATION FOR SEQ ID NO: 128: (2) SEQUENCE CHARACTERISTICS: (i) 15 LENGTH: 573 base pairs (A) TYPE: nucleic acid (B) STRANDEDNESS: single TOPOLOGY: linear (D) ORIGINAL SOURCE: (vi) ORGANISM: homosapiens (A) INDIVIDUAL ISOLATE: T2 20 SEQUENCE DESCRIPTION: SEQ ID NO: 128: (xi) ATG AGC ACA ATT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACT AAC CGT CGC CCA CAA GAC GTT AAG TTT CCG GGC 78 GGC GGC CAG ATC GTT GGC GGA GTA TAC TTG CTG CCG CGC 117 AGG GGC CCC AGG TTG GGT GTG CGC GCG ACA AGG AAG ACT 156 TCG GAG CGG TCC CAG CCT CGT GGA AGG CGC CAG CCC ATC 195 25 234 CCT AAA GAT CGG CGC TCC ACT GGC AAG TCC TGG GGA AAA CCA GGA TAC CCC TGG CCC CTG TAT GGG AAT GAG GGG CTC GGC TGG GCA GGA TGG CTC CTG TCC CCC CGA GGT TCT CGT 312 CCC TCT TGG GGC CCC AAT GAC CCC CGG CAT AGG TCG CGC 351 AAT GTG GGT AAA GTC ATC GAT ACC CTA ACG TGC GGC TTT 390 GCC GAC CTC ATG GGG TAC ATC CCC GTC GTA GGC GCC CCG 429 CTT GGT GGT GTC GCC AGA GCT CTT GCG CAT GGC GTG AGA 468 30 507 GTC CTG GAG GAC GGA GTT AAT TAT GCA ACA GGT AAC TTA CCC GGT TGC TCC TTT TCT ATC TTC TTG CTA GCC CTG CTG 546 TCC TGC ATC ACT ATT CCG GTT TCA GCT 573

(2) INFORMATION FOR SEQ ID NO: 129:

0			
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
5	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: T8	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 129:	
10	AAC ACA AAC CGC GGC GGC CAG ATC AGG GGC CCT AGG TCC GAG CGA TCC CCG AAA GAT CGG CCA GGA TAT CCT	CCT AAA CCT CAA AGA AAA ACC AAA AGA CGC CCA CAG GAC GTC AAG TTC CCG GGT GTT GGC GGA GTT TAC TTG CTG CCG CGC TTG GGT GTG CGC GCG ACA AGG AAG ACT CAG CCG CGT GGG AGA CGC CAG CCC ATC CGC TCC ACC GGC AAG TCC TGG GGA AAA TGG CCT CTT TAC GGA AAC GAG GGC TGC	39 78 117 156 195 234 273
15	CCT ACT TGG GGC AAT TTG GGC AGA GCC GAC CTC ATG GTC GGA GGC GTC CTG GAA GAC CCT GGT TGC TCT	TGG CTC CTG TCC CCC GG GGG TCT CGT CCC ACT GAC CCC GGG GAT AGA TCA CGT GTC ATC GAT ACC ATT ACA TGT GGT TTT GGG TAC ATC CCT GTC GTT GGC GCC CCG GCC AGA GCT CTG GCA CAT GGT GTT AGG GGG ATA AAC TAT GCA ACA GGG AAT TTG TTT TCT ATC TTC TTG CTT GCT CTT CTG GTG CCA GTG TCT GCA	312 351 390 429 468 507 546 573
20	(2) INFORMATI	ON FOR SEQ ID NO: 130:	
20	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
25	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US1	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 130:	
30	AAC ACA AAC CGC GGC GGT CAG ATC AGG GGC CCC AGG TCC GAG CGA TCC CCG AAA GAT CGG CCA GGA TAT CCT GGC TGG GCA GGT	CCT AAA CCT CAA AGA AAA ACC AAA AGA CGC CCA CAG GAC GTC AAG TTC CCG GGT TTG CTG GCG GGA CTT TAC TTG CTG CGC CGC GTG AGA AGA AGA ACT CAG CCG CGT ACA AGG AAG ACT CAG CCG CGT ACA AGG AAG ACT CGC TCC ACC GGC AAG CCC ATC CGC TCC ACC GGC AAG CGA AGG CGC TCC ACC GGC AAG CGC AGG CGC TCC TGT GCC CGC GGA AGG TCC TCG TGC CTC TGT CCC CCC CGC GGG TCT CGT TCC CCC C	39 78 117 156 195 234 273 312
35	CCT ACT TGG GGC	C CCC ACT GAC CCC CGG CAC AGA TCA CGT	351

	- 169 -	
0		
	AAC TTG GGC AAG GTC ATC GAT ACC ATT ACG TGT GGT TTT GCC GAC CTC ATG GGG TAC ATC CCT GTC GTT GGC GCC CCG GTC GGA GGC GTC GCC AGA GCT CTG GCA CAC GGT GTT AGG GTC CTG GAA GAC GGG ATA AAT TAC GCA ACA GGG AAT CTG CCT GGT TGC TCC TTT TCT ATC TTC TTA CTT GCT CTT CTG TGC GCC ACG GTG CCG GTG TCT GCA	390 429 468 507 546 573
5	(2) INFORMATION FOR SEQ ID NO: 131:	
10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: DK11	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:	
15 20	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAT ACA AAC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGT GGC GGC CAG AGC GGT TAG TT TAC TTG CTG CCG CGC AGG GGC GGT TT TAC TTG CTG CCC GGC AGG GGC GTT TAC TTG CTG CCC CGC AGC GGC CGC CGC CGC CGC CGC CG	39 78 117 156 195 234 273 312 351 390 429 468
25	GTC GGA GGC GTC GCC AGA GCT CTG GCA CAC GGT GTT AGA GTC CTG GAA GAC GGG ATA AAT TAC GCA ACA GGG AAT CTG CCT GGT TGC TCT TTT TCT ATC TTC TTA CTT GCT CTT CTG TCA TGC TGC ACA GTG CCA GTG TCT GCG	507 546 573
	(2) INFORMATION FOR SEQ ID NO: 132:	
30	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens	
35	(C) INDIVIDUAL ISOLATE: SW3  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:	
-		

372577\_1

۰		
5	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAT ACA AAC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGT GGC GGC CAC CAG GAC GTT AAC TTG CTG CCG CGC AGG GGC CCA CAG GTT GGC GGC ACA AGG AAG ACC ATC CCG AAA GAT CCC CGC CGC CGC CGC AGG CCC ACC CGC CGC	39 78 117 156 195 234 273 312 351 390 428 507 573
	TCG TGC TTC ACA GTG CCA GTG TCT GCG	5/3
	(2) INFORMATION FOR SEQ ID NO: 133:	
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens	
	(C) INDIVIDUAL ISOLATE: DK8	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:	
25	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACA AAC CGC CG	39 78 117 156 195 234 273 312 351 390 429 468 507
30	CCT GGT TGC TCT TTT TCT ATC TTC TTG CTT GCT CTT CTG TCG TGC TGC ACA GTG CCA GTG TCT GCG	546 573

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 573 base pairs

۰		
	(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
5	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S83	
,	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:	
10	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACT AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC GGT GGC CAG ATC GTT GGC GGA GTA TAC TTG CTG CGC CGC AGG GGC CCG AGA TTG GTT GGC GGC GGC AGG AGG AAA ACT TCC GAA CGG TCC CAG CCA CGT GGC AGG AGG AAA ACT TCC GAA GAT CGC CGC ACC ACT GGC AAG TCC TGG GAA GGC CCA ATC CCT AAA GAT CGG CGC ACC ACT GGC AAG TCC TGG GGA AGG CCC AGC CCT TGT TAT GGG AAT GAG GGC CTC GGC TGG GAC AGC CCC CGC GGT TCT CGC CCC TCC TTA TGG CCC CTC TCA TGG GGC CCC CCC TCC TCA TGC CCC CGC GGT TCT CGC CCC TCC TCA TGG GGC CCC ACC GAC CCC CCG CAT AAA TCG CGC ACC TTG GGT AAC TTG GGT AAG TCC TTC TCT CGC CCC TTC TCA TGG GGT TCT CGC CCC TTC TCA TGG GGT ACC GTC TCC TCC TCC TCC TCC TCC TCC TCC T	39 78 117 156 195 234 273 312 351 390
15	GCC GAC CTC ATG GGG TAC ATA CCC GTC GTT GGC GCT CCC GTT GGC GGC GTT GCC AGA GCC CTC GCC CAT GGG GTG AGG GTT CTG GAG GAC GGG ATA AAT TAT GCA ACG GGG AAT TTG CCC GGT TGC TCT TTC TCT ATC TTT CTC TTC GCC CTC TTG TCT TGC ATC TCT GTG CCA GTT TCC GCC	429 468 507 546 573
	(2) INFORMATION FOR SEQ ID NO: 135:	
20	(2) INFORMATION FOR SEQ ID NO: 135:  (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK10	39 117 156 195 234 273 312 351 3929

GCC CTT GAA GAC GGG ATA AAT TTC GCA ACA GGG AAC TTG CCC GGT TGC TCT TTT TCT ATC TTC CTT CTT GCT CTG TTC TCT TGC TTG TTC TCT TGC TTA ATT CAT CCA GCA GCT AGT   (i) SEQUENCE CHARACTERISTICS: (a) LENGTH: 573 base pairs (b) TYPE: nucleic acid (c) STRANDENNESS: single (D) TOPOLOGY: linear  (c) INDIVIDUAL ISOLATE: S52  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:  ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACC ATC CGT CGC CCA CAG GAC GTT AAG TTC CGC GGT AGG GGC CAA CAG TTG GGT GGA GATA TAC GTG TTG CGC GGC AGG GGA CAG ATC GGT GGT GGG CGG ACG CGG ACG CCT AAA ACT 1 TCT GAA CGG TCA CAG CCT CGC GGA CGG CGG ACG CCT ACC GGG TGC CCC AGG GAC GGA CGC TCC CGC GGG TGG CGC CTC CGC TGG CGC CCC CGG CGG	•			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  10 (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S52  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:  ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA ACA ACC ATC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGT  AGG GGA CAG ATG GTT GGT GGA GTA TAC GTG TTG CCG GGC AGG GGC CAA CAG TTG GGT GGG CGG AGG CGT AAA ACT 1 TCT GAA CGG TCA CAG CCT CGC GGA AGG ACG CCT ATC 1 TCT GAA CGG TCA CAG CCT CGC GGA AGG ACG CCT ATC 1 CCC AAG GGC CT CGG AGC GAA GGC CGG AGG CCT ACC CCC GGG TCC CGG GGA CGG CGG AGG CCT CGC GGA AGG ACG CCT CGC GGA CGC CCC AGG GGA CGG CGG AGG CCT ACC CCC GGG TCC CGG GGA CGC CCC CGG AGG CCC CCC CGG AGG CCC CCC		GCC CTT GAA GAC CCC GGT TGC TCC	GGG ATA AAT TTC GCA ACA GGG AAC TTG TTT TCT ATC TTC CTT CTT GCT CTG TTC	468 507 546 573
(A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  10 (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S52  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:  ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACC ATC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGT AGA GGC CCA CGA TTG GGT GGA GTA TAC GTG TTG CCG GGC AGA GGC CCA CGA TTG GGT GGA GTA TAC GTG TTG CCG GGC TCT GAA CGG TCA CAG CCT CGC GGA CG CAG CAG CAG ACT CCC AAG GGC CGT CGG AGC CGA CG CT AAA CCC CAG GGT CGG AGC CGA CG CT CAG GGG TGG GCA GGG TGC CTC TAT GGT AAG GGC CCC AGC CCC AGG TGG CCC AAC CAC TGG CCC CTC TAT AGT AAT GAG GGC TGC CCA TCT TGG GGC CCA AAC ACC CCC GG CGA CGC TCC GGT AAA ACT ATC GAT ACC CTT ACG TGC GGA TCC ATT TGG GGC CCA AAC ACC CCC GGC GGA GGC TCC AAT TTG GGT AAA GTC ATC GAT ACC CTT ACG TGC GGA TCC GCC GAC CTC ATG GGG TAC ATC CCT CCG CGC GGA GTC CCC GGT TAC GGC CCC AAC ACC CCC GGC CGA AGG TCC CCC GGT TCG GGC CTC CTC TAT GGT AGC TCC CGC GCC TAC TCT TGC GCC CTC TCC TCC CGC CGC AGC CCC GGT TCC TCT CTC TCT CTC TCT CTC TCC CCC GCC TCC TCC AGC ACC CCC CGC CGC AGC CCC CGT TGC TCC TTC TCT TCT TCT TCT TCT	5	(2) INFORMATIO	ON FOR SEQ ID NO: 136:	
(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S52  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:  ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACC ATC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGT GGC GGA CAG ATC GTT GGT GGA GTA TAC GTG TTG CCG CGC 1 TCT GAA CGT TCA CAG CCT CGC GGA CGA CGC AAA ACT 1 CCC AAG GCG CCT CGG AGC GGA CGA CGC TAC ATC 1 CCC AAG GCG CCT CGG AGC GAA GGC CGT CAG CCC GGG TGC CCC AGA CGC CCT CGC GGA CGG TCC CGC GGG TGC CCC AAAC GAC CCT CGC GGA CGC TCC CGC GGG TGC CCC AAAC GAC CCT CGC GGA CGC TCC CGC GGG TGG CCC AAAC GAC CCT CTG TCC CCA CGC GGC TCC CGT CCA TCT TGG GGC CCA AAC GAC CCC TAC GGC TCC CGC GGC TAC TTG GGC CCC TCT TCC CCA CGC GGC TCC CGC GGC TAC TTC GG CCC CTC TTC TCC CCA CGC GGC TCC CGC GGC TAC TTC GG CCC CTC TCC CCA CGC GGC TCC CCC GGC TAC TTC GGC CCC TC TCC CCC CGC GGC TCC CCC GGC TCC TTT GG GGC TCC TCC TCC TCC CCC C		(i)	(A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	
ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA AAC ACC ATC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGT  AGG CGA CAG ATC GTT GGT GGA GTA TAC GTG TTG CCG CGC  AGG GGC CCA CGA TTG GGT GGA GTA TAC GTG TTG CCG CGC  TCT GAA CGG TCA CAG CCT CGC GGA CGC ACG CGT AAA ACT  CCC AAG GCG CGT CGG AGC GAA GGA CGA CAG CCT ATC  CCC AGG GCG CGT CGG AGC GAA GGA CGA CAG CCT ATC  CCC GGG TAC CCT TTG CCC CTC TAT GGT TAG GGC TCC  GGG TGG GCA GGG TCG CTC CTG TCC CGA CGC GCT CCG  CCA TCT TGG GGC CCC AAC AGA CCC CGG CGA AGG TCC  CCA TCT TGG GGC CCA AAC AAC CCC CGG CGA AGG TCC  GCC GAC TCT TGG GGC CCA AAC AAC CCC CGG CGA AGG TCC  GCC GAC CTC ATG GGG TAC CCT CTG TCC CTA CGC GGC TCC  AAT TTG GGT AAA GTC ATC GAT ACC CTT ACG TGC GGA TTC  GCC GAC CTC ATG GGG TAC CTC CTC TCT CTC GTC GGC GCT  GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GGT AGG  CCC GAT CTT TCT ATT TTT TTT TTT TTT TTT TTT T	10	(vi)	(A) ORGANISM: homosapiens	
AAC ACC ATC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGT  AGG CGA CAG ATC GTT GGT GGA GTA TAC GTT GCG CGC  AGG GGC CCA CGA TTG GGT GGC GCG ACG CGT AAA ACT  TCT GAA CGG TCA CAG CCT CGC GGA CGA CGC TAAA ACT  CCC AAG GGC CCT CGG AGC GAA GGA CGA CAG CCT ATC  CCC AGG GCC CGT CGA AGG ACG CGT CGT GG GCT CAG  CCC GGG TAC CCT TGG CCC CTC TAT GGT AAT GAG GGC CTG  CCA TCT TGG GGC CCT CGC GGA CGC AGC CGT CCTGG  CCA TCT TGG GGC CCA AAC AAC CCC CGG CGG AGG TCC  GGG TGG GGA GGG TGG CTC CTG TCC CCA CGC GGC TCC CGC  AAT TTG GGT AAA GTC ATC GAT ACC CTT ACG TGC GGA TCC  GCC GAC CTC ATG GGG TAC ACC CCC GGC GGA AGG TCC  GCC GAC CTC ATG GGG TAC ATC CGC TC GTC GTC GGC GTA GGC  GCC CTT GAA GAC GGG ATA AAT TTT GCA ACA GGC GTT AGG  GCC CTT GAA GAC GGG ATA AAT TTT GCA ACA GGC GTT AGG  CCC GGT TGC TCC TTT TCT ATC TTC TCT CTT CTT GTT CTT CTT CTT CTT C		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 136:	
CCC AAG GCG CCT CGG AGC GAA GGC CGG TCC TGG GCT CAG	15	AAC ACC ATC CGT GGC GGA CAG ATC AGG GGC CCA CGA	CGC CCA CAG GAC GTT AAG TTC CCG GGT GTT GGT GGA GTA TAC GTG TTG CCG CGC TTG GGT GTG CGC GCG ACG CGT AAA ACT	39 78 117 156 195
GTA GGA GGC GTC GCA AGA GCC CTC GGG CAT GGC GTG AGG GCC CTT GAA GAG GGG ATA AAT TTT GCA ACA GGG AAC TTG CCC GGT TGC TCC TTT TCT ATC TTC CTT CTT GCT CTG TTC TCC TGC TTA GTT CAT CCT GCA GCT AGT   25  (2) INFORMATION FOR SEQ ID NO: 137:  (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  30  (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S2  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137: ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA	20	CCC AAG GCG CGT CCC GGG TAC CCT GGG TGG GCA GGG CCA TCT TGG GGC AAT TTG GGT AAA	CGG AGC GAA GGC CGG TCC TGG GCT CAG TGG CCC CTC TAT GGT AAT GAG GGC TGC TGG CTC CTG TCC CCA CGC GGC TCC CGT CCA AAC GAC CCC CGG CGG AGC TCC CGC GTC ATC GAT ACC CTT ACG TGC GGA TTC	234 273 312 351 390 429
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  30 (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S2  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:  ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA		GTA GGA GGC GTC GCC CTT GAA GAC CCC GGT TGC TCC	GCA AGA GCC CTC GCG CAT GGC GTG AGG GGG ATA AAT TTT GCA ACA GGG AAC TTG TTT TCT ATC TTC CTT CTT GCT CTG TTC	468 507 546 573
(A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  30 (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S2  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:  ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA	25	(2) INFORMATION	ON FOR SEQ ID NO: 137:	
(XI) ORIGINAL SOURCE:  (A) ORGANISM: homosapiens  (C) INDIVIDUAL ISOLATE: S2  (XI) SEQUENCE DESCRIPTION: SEQ ID NO: 137:  ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA		(i)	(A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	
ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA	30	(vi)	(A) ORGANISM: homosapiens	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 137:	
3.3	35	ATG AGC ACA CTT	CCT AAA CCT CAA AGA AAA ACC AAA AGA	3 :

372577\_1

٥		
5	AAC ACC ATC CGT CGC CCA CAG GAC ATC AAG TTC CGG GGT GGC GGA CAG ATC GTT GGT GGA GTA TAC GTG TTG CCG CGC AGG GGC CCA CAGA TTG GGT GGT GGC GCG ACG CGT AAA ACT TCT GAA CGG TCA CAG CCT CGC GGA CGG CGA CAG CCT ATC CCC AGG GCG CGA CAG CCT ATC CCC AGG GTA CCT TGG CCC CTC TAT GGT AAC GAG GCC TGC GGG TAC CCT TGG CCC CTC TAT GGT AAC GAG GGC TGC CGG TAC CCT TGG CCC CTC TAT GGT AAC GAG GGC TGC CAG TCC TTG GGC CCA TCT TGG GCC CAT CCC TCC TTG TCC CGC CGC CGC TCC CGT CCCA TCC TGG GCC CAA TCC CGC CGC CGC CGC CCC CCC TCC TCC CCC C	78 117 156 195 234 273 312 351 390 429 468 507 546 573
10	(2) INFORMATION FOR SEQ ID NO: 138:	
15	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear  (vi) ORIGINAL SOURCE:	
	(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: DK12	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:	
	ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
25	AAC ACC ATC GGT GGC CCA CAG GAC GTC AAG TTC CCG GGT GGC GGA CAG ATC GTT GGT GGA GTA TAC GTG TTT GCG CGC AGG GGC CCA CGA TTG GGT GTG GGC GGA CAG CCT AAA ACT TCT GAA CGG TCA CAG CCT CGC GGA CGG CGA CAG CCT ATC CCC AAG GCG CGT CGG AGC GAA GGC CGG ACA CAG CCT ATC CCC AGG TAC CCT TGG CCC CTC TAT GGT AAC GAG GGC TGC GGG TGG GCA GGG TGG CTC CTG TCC CCA CGC GGC TCC CGT CCA TCT TGG GGC CCA AAC GAC CCC CGG CGG AGG TCC CGT CCA TCT TGG GGC CCA AAC GAC CCC CGG CGG AGG TCC CGT AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC GGA TTC GCC GAC CTC ATG GGG TAC ATC CGC CTC CTC GTC GGC GAC TCC GTC GGG GGC GTC CCA GGG CGC CTC CTC GGC GAC CTC CTC ATG GGG TAC ATC CGC CTC CTC GTC GGC GCT CCT GTA GGG GGC GTC CCA AGG CCC CTC CTC GCG CAT GGC CTC GTA GGG GGC GTC CCA AGG GCC CTC CTC GCG CAT GGC GTG AGG	117 156 195 234 273 312 351 390 429 468
30	GCC CTT GAA GAC GGG ATA AAT TTC GCA ACA GGG AAC TTG CCC GGT TGC TCC TTT TCT ATC TTC CTT CTT GCT CTG TTC TCT TGC CTA ATT CAT CCA GCA GCT AGT  (2) INFORMATION FOR SEQ ID NO: 139:	507 546 573

(i) SEQUENCE CHARACTERISTICS: LENGTH: 573 base pairs TYPE: nucleic acid STRANDEDNESS: single (A)

(B) (C)

372577\_1

(D) TOPOLOGY: linear ORIGINAL SOURCE: (vi) (A) ORGANISM: homosapiens INDIVIDUAL ISOLATE: SEQUENCE DESCRIPTION: SEO ID NO: 139: (xi) 5 ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGC CGC CCC ATG GAC GTA AAG TTC CCG GGT 78 GGT GGC CAG ATC GTT GGC GGA GTT TAC TTG TTG CCG CGC 117 AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT CGA AAG ACT 156 TCG GAG CGG TCG CAA CCT CGT GGC AGG CGT CAA CCT ATC 195 CCC AAG GCG CGC CAG CCA GAG GGC AGA TCC TGG GCG CAG 234 CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TGC 273 10 GGG TGG GCA GGG TGG CTC CTG TCT CCT CGC GGC TCT CGG 312 CCA TCT TGG GGC CCA AAT GAT CCC CGG CGG AGA TCG CGC 351 AAT CTG GGT AAG GTC ATC GAT ACC CTG ACG TGC GGC TTC 390 GCC GAC CTC ATG GGA TAC ATC CCG ATC GTG GGC GCC CCC 429 GTG GGG GGC GTC GCC AGG GCT CTG GCG CAT GGC GTC AGG 468 GCT GTG GAG GAC GGG ATT AAC TAT GCA ACA GGG AAT CTT 507 CCC GGT TGC TCT TTC TCT ATC TTC CTT TTG GCA CTT CTT 546 15 TCG TGC CTC ACT GTT CCA GCG TCG GCT 573 (2) INFORMATION FOR SEO ID NO: 140: (i) SEQUENCE CHARACTERISTICS: LENGTH: 573 base pairs (A) (B) TYPE: nucleic acid 20 (C) STRANDEDNESS: single (D) TOPOLOGY: linear (vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: 25 SEQUENCE DESCRIPTION: SEQ ID NO: 140: (xi) ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGC CGC CCT ATG GAT GTA AAA TTC CCA GGC 78 GGC GGC CAG ATC GTT GGC GGA GTT TAC TTG TTG CCG CGC 117 AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT CGG AAG ACT 156 TCG GAG CGG TCG CAA CCT CGT GGC AGG CGT CAG CCT ATC 195 CCC AAG GCA CGT CGG TCC GAG GGT AGG TCC TGG GCT CAG 234 30 CCC GGG TAC CCA TGG CCT CTT TAC GGT AAT GAA GGC TGT 273 GGG TGG GCA GGT TGG CTC CTG TCC CCC CGC GGC TCT CGA 312 CCG TCT TGG GGC CCA AAT GAT CCC CGG CGG AGG TCG CGC 351 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC 390 GCC GAC CTC ATG GGA TAC ATC CCG CTC GTG GGC GCC CCA 429 GTA GGA GGC GTC GCC AGA GCC CTG GCG CAT GGC GTC AGG 468 GCT GTG GAG GAC GGG ATC AAC TAT GCA ACA GGG AAC CTT 507 35

0		
	CCT GGT TGC TCT TTC TCT ATC TTC CTC TTG GCA CTT CTC TCG TGC CTA ACC GTC CCA GCG TCT GCT	546 573
	(2) INFORMATION FOR SEQ ID NO: 141:	
5	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
10	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: Z1	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:	
15 20 25	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGT CGC CCC ATG GAT GTG AAA TTC CCG GGC GGC GGC GGC CAC GGC GGC GGC CAC GGC GG	39 117 156 195 234 273 3151 390 429 468 5046 573
	(C) STRANDEDNESS: single (D) TOPOLOGY: linear (vi) ORIGINAL SOURCE:	
30	(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: Z5	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:	
35	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGC CGC CCC ATG GAT GTA AAA TTC CCG GGT GGT GGT CAG ATC GTT GGC GGA GTT TAC TTG TTG CCC CGC AGG GGC CCC AGG TTG GTG CGC GCG ACT CGG AAC ACT	39 78 117 156

372577\_1

		- 176 -	
5	CCC CAG GCA CGT CCC GGG TAC CCT GGG TGG GCA GGG CCA TCT TGG GGC AAT CTG GGT AAG GCC GAC CTC ATG GTA GGT GGC GTC	CAA CCT CGC GGC AGG CGT CAG CCT ATC CGG TCC GAG GGC AGG TCC TGG GCT CAG TGG CCT CTT TAT GGC AAT GAG GGC TGT TGG CTC CTG TCC CCC CGC GGA TCT CGG CAA AAT GAT CCC CGG CGT AGG TCC CGC GTC ATC ATC ACC CTG AGG TGT GGC TTC GGA TAC ATT CCG CTC GTC GGC GCC CCA GCC AGG GCC TTG GGC CAT GGC GTC AGG GGA ATC AAC TAT GCA ACA GGG AAT CTT	195 234 273 312 351 390 429 468 507
	CCT GGT TGC TCC TCG TGC TTG ACA	TTT TCT ATC TTC CTA CTT GCA CTT TTC ACA CCG GCA TCC GCT ON FOR SEQ ID NO: 143:	546 573
10	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
15	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: Z6	
20	ACC ACC CAC GCG GGT GGC GGC GGC GGC GGC GGC GGC GG	SEQUENCE DESCRIPTION: SEQ ID NO: 143:  CCT AAA CCT CAA AGA AAA ACC AAA CGT CGC CCC ATG GAC GTT AAG TTC CCG GGT GTT GGC GGA GTT TAC TTG TTG CCG CGC TTG GGT GTG CGC GCG ACT AGG AAG ACT CAA CCT CGT GGG AGA CGC CAG CCT ATC CGA TCT GAG GGA AGG TCC TGG GCT CTG TGG CTC CTT TAC CCC CGT GGC TCT CGA CCC AAT GAT CCC CGG CGA AGG TCC CCA AAT GAT CCC CGG CGA CCC CGC CCC AGG CCC CTG TCG GGC TCT TCC CGC CGC GGA AGG TCC CCC AGT ACC CCC CTG TCC TTC GGA TAC ATC CCC CTG TGA GCC CCC GCC AGG GCC CTG GCA AGG TCT CTC TCC TCT TCC CTC TTG GCA CTC TTC CGG TCC AAT TCC CCC CTC TTG GCC CCC CCC AGG GCC CTG GCA CAT GGT TTT TTC TCT ATC TTC CTC TTG GCA CTT CTT CTT CCC ACC TCG GCC	39 78 117 156 195 234 273 312 351 390 429 468 507 546
30	(2) INFORMATI	ON FOR SEQ ID NO: 144:  SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
35	(vi)	ORIGINAL SOURCE:	

372577\_1

```
(A)
                          ORGANISM: homosapiens
                      (C)
                          INDIVIDUAL ISOLATE: Z7
            (xi)
                     SEQUENCE DESCRIPTION: SEQ ID NO: 144:
     ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT
                                                                39
                                                                78
     AAC ACC AAC CGC CGC CCC ATG GAC GTT AAG TTC CCG GGC
     GGT GGC CAG ATC GTT GGC GGA GTT TAC TTG TTG CCG CGC
                                                               117
     AGG GGC CCC AGA TTG GGT GTG CGC ACA ACT AGG AAG ACT
     TCG GAG CGG TCG CAA CCT CGT GGG AGA CGT CAG CCT ATC
                                                               195
     CCC AAG GCA CGT CGA TCT GAG GGA AGG TCC TGG GCT CAA
                                                               234
     CCC GGG TAC CCA TGG CCT CTT TAC GGT AAC GAG GGT TGC
                                                               273
     GGG TGG GCA GGA TGG CTC TTG TCA CCC CGT GGC TCT CGA
                                                               312
     CCG TCT TGG GGC CCA AAT GAT CCC CGG CGA AGG TCC CGC
                                                               351
     AAC TTG GGT AAG GTC ATC GAT ACC CTA ACC TGC GGC TTT
                                                               390
10
     GCC GAC CTC ATG GGA TAC ATC CCG CTC GTA GGC GCC CCC
                                                               429
     GTG GGC GGC GTC GCC AGG GCC CTA GCG CAT GGC GTT AGG
                                                               468
     GCT CTG GAG GAC GGG ATT AAT TAT GCA ACA GGG AAC CTT
                                                               507
     CCC GGT TGC TCT TTT TCT ATC TTC CTC TTG GCA CTT CTT
                                                               546
     TCG TGC CTG ACT GTT CCC GCC TCG GCC
                                                               573
15
           INFORMATION FOR SEO ID NO: 145:
     (2)
             (i)
                     SEQUENCE CHARACTERISTICS:
                      (A)
                          LENGTH: 573 base pairs
                      (B)
                           TYPE: nucleic acid
                          STRANDEDNESS: single
                      (D)
                          TOPOLOGY:
                                      linear
20
             (vi)
                     ORIGINAL SOURCE:
                      (A)
                          ORGANISM: homosapiens
                           INDIVIDUAL ISOLATE: DK13
                      (C)
             (xi)
                     SEQUENCE DESCRIPTION: SEQ ID NO: 145:
     ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT
                                                                 39
25
     AAC ACC AAC CGC CGC CCA ATG GAC GTT AAG TTC CCG GGT
                                                                78
     GGC GGC CAG ATC GTT GGC GGA GTT TAC TTG TTG CCG CGC
                                                               117
     AGG GGC CCT AGA TTG GGT GTG CGC GCG ACT AGG AAG ACT
                                                               156
     TCG GAG CGG TCG CAA CCT CGT GGG AGG CGC CAG CCT ATC
                                                               195
     CCC AAG GCG CGC CAA CTC GAG GGT AGG TCC TGG GCT CAG
                                                               234
     CCT GGG TAT CCT TGG CCC CTT TAC GGC AAT GAG GGC TGC
                                                               273
     GGG TGG GCG GGA TGG CTC CTG TCA CCC CGT GGC TCT CGG
                                                               312
     CCG TCT TGG GGC CCG AAT GAT CCC CGG CGG AGG TCC CGC
                                                               351
30
     AAC TTG GGT AAG GTC ATC GAT ACC CTA ACT TGC GGC TTC
                                                               390
     GCC GAC CTC ATG GGA TAC ATC CCG GTC GTA GGC GCC CCC
                                                               429
     GTG GGT GGC GTC GCC AGA GCC CTG GCG CAT GGC GTC AGG
                                                               468
     CTT CTG GAG GAC GGG GTC AAT TAT GCA ACA GGG AAT CTT
                                                               507
     CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCA CTG CTC
                                                               546
     TCG TGC CTG ACT GTT CCC GCT TCG GCC
                                                                573
```

c

372577\_1

	(2) INFORMATION FOR SEQ ID NO: 146:										
5	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear										
	(vi) ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA4										
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 146:									
10	AAC ACC AAC CGC	CCT AAA CCT CAA AGA AAA ACC AAA AGA CGC CCA CAG GAC GTT AAG TTC CCG GGC	39 78								
15	AGG GGC CCT AGG TCA GAA CGG TCG CCC AAG GCG CGC CCC GGG TAC CCT AAT TGG GGC AAT TTG GGT AAG GCC GAC CTC ATG GTT GGG GGC GTC CTT TTT GAG GAC CCC GGT TGC TCT TCT CCC GGT TGC TCC	GTT GGT GGA GTC TAC TTG TTG CCG CGC TTG GGT GTG CGC GCG ACT CGG AAG ACT CAA CCC CGT GGG CGG CGC CAG CCT ATT CAA CCC ACG GGC CGG CGC CAG CCT ATT GG CCC CTT TAC GCC ACT GAG GGC CTC TGG TTG CTC TCC CCC CGA GGC TCT CGG CCC AAT GAC CCC CGG CGA AAG TCG CGC GTC ATC GAT ACC CTA ACG TGC GGA TTC GGG TAC ATC CCG CTA ACG TGC GGA TTC GGG TAC ATC CCG CTC GTA GGC GGC CCC CCA AGG GCC CTT GCA CAT GGT GTG AGG GGG GTA AAC TAT GCA CAT GGT GTG AGG GGG GTA AAC TAT GCA CAT GGT GTG AGG GGG GTA CAC TTT TATC CTT GCA CTT CTC GTC CCG GCC TCT GCA	117 156 195 234 273 312 351 390 468 507 546 573								
	(2) INFORMATI	ON FOR SEQ ID NO: 147:									
25	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear									
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA5									
30	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 147:									
35	AAC ACC AAC CGC GGT GGT CAG ATC AGG GGC CCT AGA TCA GAA CGG TCG	CCT AAA CCT CAA AGA AAA ACC AAA AGA CGC CCA CAG GAC GTC AAG TTC CCG GGC GTT GGT GGA GTT TAC TTG TTG CCG CGC TTG GGT GTG CGC GCG ACT CGG AAG ACT CAA CCC CGT GGG CGG CGC CAG CCT ATT CAA CCC ACG GGC CGG TCC TGG GGT CAA	39 78 117 156 195 234								

		- 179 -						
0								
5	GGG TGG GCA GGG CCT AAT TGG GGC AAT TTG GGT AAG GCC GAC CTC ATG GTT GGG GGC GTC GTT CTT GAG GAC CCC GGT TGC TCT	TGG	73 12 51 90 29 68 07 46 73					
	(2) INFORMATI	ON FOR SEQ ID NO: 148:						
10	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear						
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA7						
15	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 148:						
20	AAC ACC AAC GCC GGT GGT CAG ATC AGG GGC CCT AGG TCA GAA GGG TCG CCC AAG GGG CGC CCC GGG TAC CCT AAT TGG GGC AAT TTG GGT AAG GCC GAC CTC ATG GTT GGG GTC GTC TCTT GAG GAC CCC GGT TGC TCC CCC GGT TGC TCC	GGC   CGA   CAG   GAC   GTC   AAG   TTC   CGG   GGC	39 78 77 56 95 12 12 90 90 46 73					
	(2) INFORMATI	ON FOR SEQ ID NO: 149:						
30	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear						
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA1						

```
(xi)
                     SEOUENCE DESCRIPTION: SEO ID NO: 149:
     ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA
                                                                39
     AAC ACC AAC CTC CGC CCA CAG GAC GTC AAG TTC CCG GGC
                                                                78
     GGT GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC
                                                               117
     AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT CGG AAG ACT
                                                               156
     TCG GAA CGG TCG CAA CCC CGT GGG CGG CGC CAG CCT ATT
5
     CCC AAG GCG CGC CAA CCC ACG GGC CGG TCC TGG GGT CAA
     CCC GGG TAC CCT TGG CCC CTT TAC GCC AAT GAG GGC CTC
                                                               273
     GGG TGG GCA GGG TGG TTG CTC TCC CCC CGA GGC TCT CGG
                                                               312
     CCT AAT TGG GGC CCC AAT GAC CCC CGG CGG AAG TCG CGC
                                                               351
     AAT TTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGA TTC
                                                               390
     GCC GAC CTC ATG GGG TAC ATC CCG CTC GTA GGC GGC CCC
                                                               429
     GTT GGG GGC GTC GCA AGG GCT CTC GCA CAC GGT GTG AGG
                                                               468
     GTT CTT GAG GAC GGG GTA AAC TAC GCA ACA GGG AAT TTG
                                                               507
10
     CCC GGT TGC TCT TTC TCT ATC TTT ATC CTT GCA CTT CTT
                                                               546
     TCC TGT CTG ATC ATC CCG GCC TCT GCA
                                                               573
           INFORMATION FOR SEQ ID NO: 150:
     (2)
            (i)
                     SEQUENCE CHARACTERISTICS:
15
                      (A)
                          LENGTH:
                                  573 base pairs
                      (B)
                          TYPE: nucleic acid
                      (C)
                           STRANDEDNESS: single
                      (D)
                          TOPOLOGY:
                                      linear
            (vi)
                     ORIGINAL SOURCE:
                     (A)
                          ORGANISM: homosapiens
                     (C)
                           INDIVIDUAL ISOLATE:
20
            (xi)
                     SEQUENCE DESCRIPTION: SEQ ID NO: 150:
     ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA
     AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC
                                                                78
     GGT GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC
                                                               117
     AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT CGG AAG ACT
                                                               156
25
     TCA GAA CGG TCG CAA CCC CGT GGA CGG CGC CAG CCT ATT
                                                               195
     CCC AAG GCT CGC CAG CCC ACG GGC CGG TCC TGG GGT CAA
                                                               234
     CCC GGG TAC CCT TGG CCC CTT TAC GCC AAT GAG GGC CTC
                                                               273
     GAG TGG GCA GGG TGG TTG CTC TCC CCC CGA GGC TCT CGG
                                                               312
     CCT AGT TGG GGC CCC AAC GAC CCC CGG CGG AAA TCG CGC
                                                               351
     AAT TTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGA TTC
                                                               390
     GCC GAT CTC ATG GGG TAC ATC CCG CTC GTA GGC GGC CCC
                                                               429
     GTT GGG GGC GTC GCA AGG GCT CTC GCA CAT GGT GTG AGG
                                                               468
30
     GTT CTT GAG GAC GGG GTA AAC TAC GCA ACA GGG AAT TTA
                                                               507
```

(2) INFORMATION FOR SEQ ID NO: 151:

TCA TGC CTG ACC GTC CCG GCC TCT GCA

CCC GGT TGC TCT TTC TCT ATC TTT ATC CTT GCA CTT CTT

546

573

۰			
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
5	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA13	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 151:	
10	AAC ACC AAC CGC GGT GGT CAG ATC AGG GGC CCT AGG TCA GAA CGG TCG CCC AAG GCG CGC CCC GGG TAC CCT	CCT AAA CCT CAA AGA AAA ACC AAA AGA CGC CCA CAG GAC GTC AAG TTC CCG GGC GTT GGT GGA GTT TAC TTG TTG CCG CGC TTG GGT GTG CGC GCA ACT CGG AAG ACT CAA CCC CGT GGA CGG CGT CAG CCT ATC CAG CCC ACG GGC CGG TCC TGG GGT CAA TGG CCC CTT TAT GCC AAT GAG GGC CTC	39 78 117 156 195 234 273
15	CCT AAT TGG GGC AAC TTG GGT AAG GCC GAC CTC ATG GTT GGG GGC GTC GTC CTT GAG GAC CCC GGT TGC TCT	TGG TTG CTC TCC CCC CGA GGC TCT CGG CCC AAT GAC CCC CGG CGG AAA TCG CGC GTC ATC GAT ACC CTG ACG TGC GGA TTC GGG TAC ATC CCG CTC GTA GGC GGC CCC GCA AGG GCT CTC GCA CAC GGT GTG AGG GGG GTA AAC TAT GCA ACA GGG AAT TTA TTC TCT ATC TTT ATC CTT GCA CTT CTT GTC CCG ACC TCT GCC	312 351 390 429 468 507 546 573
20	(2) INFORMATION	ON FOR SEQ ID NO: 152:	
	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
25	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA6	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 152:	
30	AAC ACC AAC CGC GGT GGT CAG ATC AGG GGC CCT CGT TCG GAA CGG TCG CCC AAG GCG CGC CCC GGG TAC CCT GGG TGG GCA GGG	CCT AAA CCT CAA AGA AAA ACC CAA AGA CGC CCA CAG GAC GTC AAG TTC CCG GGC GTT GGT GGA GTT TAC TTG TTG CCG CGC ATG GGT GTG CGC GCG ACT CGG AAG ACT CAA CCC CGT GGA CGG CGT CAG CCT ATT CAA TCC GCG GGT CGG TCC TGG GGT CAA TGG CCC CTT TAC GCC AAT GAG GGC CTC TGG TTG CTC TCC CCC CGA GGC TCT CGG	39 78 117 156 195 234 273 312
35	CCT AAT TGG GGC	CCC AAT GAC CCC CGG CGA AAA TCG CGC	351

372577\_1

0		
	GCC GAC CTC ATG GTT GGG GGC GTC GTT CTT GAG GAC CCC GGT TGC TCT	GTC ATC GAT ACC CTA ACG TGC GGA TTC 399 GGG TAC ATC CCG CTC GTA GGC GGC CCC 422 GCA AGG GGT CTC AGG AGG GGG GTA AAC TAT GCA ACA GGG AAT TTG 500 TTC TCT ATC TTT GTC CTT GCA CTT CTC 544 GTC CCT GCC TCT GCA CTT CTC 547 GTC CCT GCC CTT GCA CTT CTC 547 GTC CCT GCC CTT GCA CTT CTC 547 GTC CTT GCA CTT CTC 547 GTC CCT GCC CTT GCA CTT CTC 547 GTC CCT GCC CTT GCA CTT CTC 547 GTC CTC CTC CTC CTC CTC CTC CTC CTC CTC
5	(2) INFORMATIO	ON FOR SEQ ID NO: 153:
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 573 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
10	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SA11
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 153:
15	AAC ACC AAC CGC GGT GGT CAG ATC AGG GGC CCT AGG	CCT AAA CCT CAA AGA AAA ACC AAA AGA         33           CGC CCA CAG GAC GTC AAG TTC CCG GGC         78           GTT GGT GGA GTT TAC TTG TTG CCG CGC         11'           TTG GGT GTG CGC GCG ACT CGG AAG ACT         156           CAA CCC CGT GGG CGG CGT CAG CCT ATT         199
20	CCC AAG GCG CGC CCC GGG TAC CCT GGG TGG GCA CCT AAC TGG GGC AAT TTG GGC AAG GCC GAC CTC ATG GTT GGG GGC GTC GCT CTT GAG GAC CCC GGT TGC TCT	CAA CCC ACG GGC CGG TCC TGG GGT CAA 237 TGG CCG TTT TAC GCC AAT GAG GGC CTC CCC AAT GAC CCC CGT GCA GGC TCC CGG 317 CCC AAT GAC CCC CGG GAA AGA TCC CGG GTC ATC GAT ACC CTA ACG TGC GGA TTC GGG TAC ATC CGC CTC GTA GGC GGC ACC GGG TAC ATC CGC CTC GTA GGC GGC CCC GCA AGG GCC CTC GTA GGC GGC CCC GCA AGG GCC CTC GTA GCG GGC ACC GCG AAT AAT TAT GCA ACA GGG AAT CTT TTC TCC ATC TTT ATC CTT GCA CTT CTC 544 GTC CCG GCC ACT GCA 577
25	(2) INFORMATIC	ON FOR SEQ ID NO: 154:
	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 573 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear
30	(vi)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK2
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 154:
35	ATG AGC ACA CTT	CCA AAA CCC CAA AGA AAA ACC AAA AGA 3

372577\_1

	AAC	ACC	AAC	CGT	CGC	CCA	ACG	GAC	GTC	AAG	TTC	CCG	GGT	78
	GGC	GGT	CAG	ATC	GTT	GGC	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
	AGG	GGC	CCC	CGG	TTG	GGT	GTG	CGC	GCG	ACG	AGA	AAG	ACT	156
	TCC	GAG	CGA	TCC	CAG	CCC	AGA	GGC	AGG	CGC	CAA	CCT	ATA	195
	CCA	AAG	GCG	CGC	CAG	CCC	CAG	GGC	AGG	CAC	TGG	GCT	CAG	234
	CCC	GGA	TAC	CCT	TGG	CCT	CTT	TAT	GGA	AAC	GAG	GGC	TGT	273
5	GGG	TGG	GCA	GGT	TGG	CTC	CTG	TCC	CCC	CGC	GGC	TCC	CGG	312
5	CCA	CAT	TGG	GGC	CCC	AAT	GAC	CCC	CGG	CGT	CGA	TCC	CGG	351
	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTA	ACG	TGT	GGG	TTC	390
	GCC	GAT	CTC	ATG	GGG	TAC	ATT	CCC	GTC	GTG	GGC	GCG	CCT	429
	TTG	GGC	GGC	GTC	GCG	GCT	GCG	CTC	GCA	CAT	GGC	GTG	AGG	468
	GCA	ATC	GAG	GAC	GGG	ATC	AAT	TAT	GCA	ACA	GGG	AAT	CTC	507
	CCC	GGT	TGC	TCT	TTC	TCT	ATC	TTC	CTT	TTG	GCA	CTA	CTC	546
	TCG	TGC	CTC	ACA	ACG	CCA	GCT	TCG	GCT					573
10														

## (2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

145

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: DK7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

20 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 30 35 Arg Leu Gly Val Arg Ala Pro Arg Lys Thr Ser Glu Arg Ser 25 45 50 55 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 60 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 30 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 125 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 130 135 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly

35

Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 170 175 Ser Cys Leu Thr Val Pro Ala Ser Ala 185

5

10

15

#### (2) INFORMATION FOR SEQ ID NO: 156:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - TOPOLOGY: unknown

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn

#### (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- INDIVIDUAL ISOLATE: US11

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

- 5 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 30 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 20 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 60 65 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 95 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 2.5 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 125 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 130 135 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn
- 155 160 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 180 Ser Cys Leu Thr Val Pro Ala Ser Ala

185 190

## (2) INFORMATION FOR SEQ ID NO: 157:

- (i) SEOUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: S14
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 55

Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
60
Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro
75
Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu

85 90 95 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 100 105 110

Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn

155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 170 175 180

Ser Cys Leu Thr Val Pro Ala Ser Ala 185 190

## 30 (2) INFORMATION FOR SEQ ID NO: 158:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

20

(vi)	ORIGINAL	SOURCE
( )	OTTOTTIL	DOOLLOD.

- (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: SW
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro
75
80
Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu
85
90
95
Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro

100 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 125

Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 130 135 140 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 145 150

Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 155 160 165
Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Ala Leu Leu 170 175 180

Ser Cys Leu Thr Val Pro Ala Ser Ala 185 190

(2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid
(C) STRANDEDNESS: un

(C) STRANDEDNESS: unknown(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 1 5 10

35

30

15

Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 5 60 65 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 95 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 105 110 10 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 130 135 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 155 160 165 15 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 170 175 Ser Cys Leu Thr Val Pro Ala Ser Ala 185

## (2) INFORMATION FOR SEO ID NO: 160:

(i)

20

25

30

# SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

C) INDIVIDUAL ISOLATE: DR4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 95 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 100 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 5 115 120 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 10 175 Ser Cys Leu Thr Val Pro Ala Ser Ala 185 190

(2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
- 20 (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: SA10
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 20 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 30 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 60 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro 30 75 80 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 105

Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu

120

9

372577\_1

115

35

15

20

25

30

Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 165 160 Leu Pro Gly Cys Pro Phe Ser Ile Phe Leu Leu Ala Leu Leu 180 175 170 Ser Cys Leu Thr Ile Pro Ala Ser Ala 190 185

### INFORMATION FOR SEQ ID NO: 162: (2)

SEQUENCE CHARACTERISTICS: (i) 10

LENGTH: 191 amino acids (A)

TYPE: amino acid (B)

STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

ORIGINAL SOURCE: (vi)

(xi)

ORGANISM: homosapiens (A) INDIVIDUAL ISOLATE: S45

162:

(C) SEQUENCE DESCRIPTION: SEQ ID NO:

Met Ser Thr Asn Pro Lys Pro Gln Arg Ala Thr Lys Arg Asn 5 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 15 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 65 60 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly His Pro Trp Pro 75 80 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 95 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 105 100 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 125 120 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 165 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 180

35

10

20

25

Ser Cys Leu Thr Ile Pro Ala Ser Ala 185

### INFORMATION FOR SEQ ID NO: (2)

- SEQUENCE CHARACTERISTICS: (i) 5
  - (A) LENGTH: 191 amino acids
    - TYPE: amino acid (B)
    - STRANDEDNESS: unknown (C)
    - (D) TOPOLOGY: unknown
    - ORIGINAL SOURCE: (vi)
      - ORGANISM: homosapiens (A)
      - INDIVIDUAL ISOLATE: D1 (C)
    - SEQUENCE DESCRIPTION: SEQ ID NO: 163: (xi)

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn

10 5 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro

15 35 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 55 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg

65 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro 80 75

Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 90 95 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 110 105 100

Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 125 120 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130

- Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 165 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 180 175
- Ser Cys Leu Thr Ile Pro Ala Ser Ala 30 190 185

#### INFORMATION FOR SEQ ID NO: 164: (2)

SEQUENCE CHARACTERISTICS: (i) (A) LENGTH: 191 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: US6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 15 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 40 3.5 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 65 60 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro 80 75 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu 95 90 85 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 110 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 125 120 115 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 140 135 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 165 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 180 175

## (2) INFORMATION FOR SEQ ID NO: 165:

Ser Cys Leu Thr Ile Pro Ala Ser Ala

185

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids

190

- (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: P10

35

30

5

10

15

20

						SEQUENCE DESCRIPTION: SEQ ID NO: 165:								
		(	(xi)		SEQU	ENCE	DES	SCRIE	LION	1: 2	SEQ 1	ט אכ	,: 1	.05:
	7				- 5					10	Thr			
	Thr				Pro	20					Pro 25			
5	Gln	20				Val	3.5				Arg	40		
		Leu	45					50			Ser		22	
			Arg	60					65		Lys			, ,
					75					80	Tyr			
10	0.5					90					Gly 95			
	Ser	100					105				Pro	TIU		
	_	Arg	115					120			Ile		125	
				130					135		Ile			140
15					145					T20				
	1					160					Ala 165			
	Leu	Pro		Cys	Ser	Phe	Ser 175	Ile	Phe	Leu	Leu	Ala 180	Leu	Let
20	Ser			Thr	Ile	Pro	Ala	Ser 190	Ala					

# (2) INFORMATION FOR SEQ ID NO: 166:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: DK1
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

35

20

0 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 70 65 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu 5 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 115 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 140 135 130 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 10 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 165 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 180 Ser Cys Leu Thr Ile Pro Ala Ser Ala 190 185

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids
(B) TYPE: amino acid

) TYPE: amino acid ) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens
(C) INDIVIDUAL ISOLATE: T10

(C) INDIVIDUAL ISOLATE: T10

SEQUENCE DESCRIPTION: SEQ ID NO: 167: (xi) 25 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 25 15 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 40 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 30 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 60 65 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 80 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu 85 90

Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 105 100 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 125 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 165 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 180 175 Ser Cys Leu Thr Ile Pro Ala Ser Ala 185 190

10

\_

5

### INFORMATION FOR SEQ ID NO: 168: (2)

SEQUENCE CHARACTERISTICS: (i)

191 amino acids (A) LENGTH:

amino acid TYPE:

STRANDEDNESS: unknown (C)

(D) TOPOLOGY: unknown

15

### ORIGINAL SOURCE: (vi)

145

ORGANISM: homosapiens (A)

(C) INDIVIDUAL ISOLATE:

SEQUENCE DESCRIPTION: SEQ ID NO: 168: (xi)

20 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 3.0 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 25 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 60 65 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 80 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu 95 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 30 110 105 100 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 150

0 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 165 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 180 170 175 Ser Cys Leu Thr Ile Pro Ala Ser Ala 185

5

- INFORMATION FOR SEQ ID NO: (2)
  - SEQUENCE CHARACTERISTICS: (i)
    - LENGTH: 191 amino acids (A)
    - TYPE: amino acid (B)
    - (C) STRANDEDNESS: unknown
    - TOPOLOGY: unknown (D)
  - (vi) ORIGINAL SOURCE:
    - ORGANISM: homosapiens (A)
    - INDIVIDUAL ISOLATE: IND3 (C)
  - SEQUENCE DESCRIPTION: SEQ ID NO: 169: (xi)
- 15 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 1o Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35
  - 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 55 45 50
- 20 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 70 65 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro
  - 80 75 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 90
- Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 25 110 100 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120
  - 115 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 150 145
- Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 30 160 165 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
  - 170 175 Ser Cys Leu Thr Ile Pro Ala Ser Ala 185

(2) INFORMATION FOR SEQ ID NO: 170:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: IND8
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:
- Arg Let Gly var Arg Ala Int Arg bys Int Set Gla Arg
  45
  50
  55
  15 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
- Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly His Pro Trp Pro
  75
  Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu
  85
  90
  91

Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
100 105 110
Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu

Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val

Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn
155
Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
170
175
180

Ser Cys Leu Thr Val Pro Ala Ser Ala 185 190

- 30 (2) INFORMATION FOR SEQ ID NO: 171:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 191 amino acids
    - (B) TYPE: amino acid
      - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
  - (vi) ORIGINAL SOURCE:

372577 1

35

20

٥

- ORGANISM: homosapiens (A)
- INDIVIDUAL ISOLATE: S9 (C)
- SEQUENCE DESCRIPTION: SEQ ID NO: 171: (xi)

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 5 5 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 40 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 55 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg His 60 65 10 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro 80 75 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 15 120 125 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 20 175 Ser Cys Leu Thr Ile Pro Ala Ser Ala

INFORMATION FOR SEQ ID NO: (2)

185

25 SEQUENCE CHARACTERISTICS: (i)

LENGTH: 191 amino acids

TYPE: amino acid (B) STRANDEDNESS: unknown

190

(C) TOPOLOGY: unknown (D)

ORIGINAL SOURCE: (vi)

(A) ORGANISM: homosapiens

INDIVIDUAL ISOLATE: HK3

SEQUENCE DESCRIPTION: SEQ ID NO: 172: (xi)

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 5

35

Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 65 5 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro 80 75 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu 95 90 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Thr Asp Pro 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 115 10 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 165 160 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 15 175 170 Ser Cys Leu Thr Thr Pro Ala Ser Ala 190 185

# (2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid(C) STRANDEDNESS: unknown

(C) SIRANDEDNESS: UNKNOW

(D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: HK5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

20

25

Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro 20 75 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu 95 90 Ser Pro His Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 125 120 115 5 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 140 135 130 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 Ile Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 170 10 Ser Cys Leu Thr Thr Pro Val Ser Ala 185

### INFORMATION FOR SEQ ID NO: (2)

SEQUENCE CHARACTERISTICS: (i)

191 amino acids (A) LENGTH:

TYPE: amino acid (B)

STRANDEDNESS: unknown (C)

(D) TOPOLOGY: unknown

ORIGINAL SOURCE: (vi)

ORGANISM: homosapiens (A)

INDIVIDUAL ISOLATE: HK4 20

> SEQUENCE DESCRIPTION: SEQ ID NO: 174: (xi)

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 60 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro 80 75 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 100 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 125

120

30

15

25

Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Val Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 165 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 170 5 Ser Cys Leu Thr Ile Pro Ala Ser Ala 190 185

## (2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

# (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

C) INDIVIDUAL ISOLATE: P8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

Met Ser Thr Thr Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Ser Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 40 35 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 60 65 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly His Pro Trp Pro 75 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 95 90 85 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 105 100 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 125 115 120 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 130 135 Gly Gly Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Val Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu

175

180

35

30

10

15

20

25

- Ser Cys Leu Thr Ile Pro Ala Ser Ala 185
  - INFORMATION FOR SEQ ID NO: 176: (2)
    - SEOUENCE CHARACTERISTICS:
- LENGTH: 191 amino acids (A) 5
  - TYPE: amino acid
  - STRANDEDNESS: unknown
  - TOPOLOGY: unknown (D)
  - ORIGINAL SOURCE: (vi) (A) ORGANISM: homosapiens
    - INDIVIDUAL ISOLATE:
- (C)
- 10 SEQUENCE DESCRIPTION: SEQ ID NO: 176: (xi)

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20

Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 15 35 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser

50 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 65 60 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro

75 Leu Tyr Gly Asp Glu Gly Met Gly Trp Ala Gly Trp Leu Leu 20 90

Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Thr Asp Pro 100 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 115 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val

130 135 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145

Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 170 175

Ser Cys Leu Thr Ile Pro Ala Ser Ala 185

- (2) INFORMATION FOR SEQ ID NO:
  - SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid

35

25

- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
    - (C) INDIVIDUAL ISOLATE: T4
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg Arg 60 65 70 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro 75 80

Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 85 90 95
Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro

100 105 110 110 Arg His Arg Ser Arg Asn Val Gly Lys Val Ile Asp Thr Leu 125 125 Thr Cys Ser Leu Ala Asp Leu Met Gly Tyr Val Pro Val Val 130 135 140

20 Gly Gly Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly
145 150

Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn
155 160 165

Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
170 170 180

Ser Cys Ile Thr Ile Pro Val Ser Ala 185 190

25 (2) INFORMATION FOR SEQ ID NO: 178

- (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
- 30 (D) TOPOLOGY: unknown
  - (vi) ORIGINAL SOURCE:

    (A) ORGANISM: homosapiens
    - (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: US10
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

35

0

10

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 5 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 5 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg 65 Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro 75 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 95 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 10 105 Arg His Arg Ser Arg Asn Val Gly Lys Val Ile Asp Thr Leu 120 115 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val 135 130 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 15 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 Ser Cys Ile Thr Ile Pro Val Ser Ala

20
(2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid (C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

25 (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: T9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Ser Asp Pro Arg His Arg Ser Arg Asn Val Gly Lys Val Ile Asp Thr Leu Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys Ile Thr Thr Pro Ala Ser Ala

15 (2) INFORMATION FOR SEQ ID NO: 180:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
    (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown

- D) TOPOLOGY: unknown
- 20 (vi) ORIGINAL SOURCE:
  (A) ORGANISM: homosapiens
  (C) INDIVIDUAL ISOLATE: T
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:
- Met Ser Thr Ile Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro

372577\_1

Arg His Arg Ser Arg Asn Val Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys Ile Thr Ile Pro Val Ser Ala 

# 10 (2) INFORMATION FOR SEQ ID NO: 181:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: T8
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg Gly Ser Arg Pro Thr Trp Gly Pro Thr Asp Pro Arg His Arg Ser Arg Asn Leu Gly Arg Val Ile Asp Thr Ile Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn

Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 170 175 180 Ser Cys Phe Thr Val Pro Val Ser Ala 190 185

INFORMATION FOR SEQ ID NO: 182: (2) 5

> SEQUENCE CHARACTERISTICS: (i)

191 amino acids (A) LENGTH:

TYPE: amino acid (B) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

INDIVIDUAL ISOLATE: US1

SEQ ID NO: 182: (xi) SEOUENCE DESCRIPTION:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 5 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 15 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg 60 65 20 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro 75 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Thr Trp Gly Pro Thr Asp Pro 110 100 105 Arg His Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Ile 125 25 120 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val 130 135 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 160 165 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 30

175

180

Ser Cys Ala Thr Val Pro Val Ser Ala 185 190

> INFORMATION FOR SEQ ID NO: (2)

(i) SEQUENCE CHARACTERISTICS: LENGTH: 191 amino acids (A) TYPE: amino acid (B) (C) STRANDEDNESS: unknown (D) TOPOLOGY: unknown (vi) ORIGINAL SOURCE: 5 (A) ORGANISM: homosapiens INDIVIDUAL ISOLATE: DK11 (C) SEQUENCE DESCRIPTION: SEQ ID NO: (xi) 183: Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 10 20 25 15 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 40 30 Arg Leu Gly Val Arg Thr Thr Arg Lys Thr Ser Glu Arg Ser 50 45 Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg 65 60 15 Ser Thr Gly Lys Pro Trp Gly Lys Pro Gly Tyr Pro Trp Pro 75 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 95 90 85 Ser Pro Arg Gly Ser His Pro Asn Trp Gly Pro Thr Asp Pro 105 110 Arg His Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Ile 115 120 20 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val 135 130 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 25 170 175 Ser Cys Cys Thr Val Pro Val Ser Ala 185 190 (2) INFORMATION FOR SEQ ID NO: 184:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
- (A) ORGANISM: homosapiens

35

## (C) INDIVIDUAL ISOLATE: SW3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 5 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 40 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg 65 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro 10 75 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser His Pro Asn Trp Gly Pro Thr Asp Pro 110 105 Arg His Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Ile 115 120 15 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val 135 140 130 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 160 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 20 Ser Cys Phe Thr Val Pro Val Ser Ala

(2) INFORMATION FOR SEQ ID NO: 185:

185

2) INFORMATION FOR SEQ ID NO. 105.

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: DK8

30 (xi) SEOUENCE DESCRIPTION: SEO ID NO: 185:

35

25

Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Ser Ser Glu Arg Ser 50 Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg 65 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro 80 75 5 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Thr Trp Gly Pro Thr Asp Pro 105 Arg His Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Ile 120 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val 130 135 10 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 Ser Cys Cys Thr Val Pro Val Ser Ala 15 185 190

### (2) INFORMATION FOR SEO ID NO: 186:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: S83

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

 Met
 Ser
 Thr
 Asn
 Pro
 Lys
 Pro
 Gln
 Arg
 Lys
 Thr
 Lys
 Arg
 Asn
 Asn
 Arg
 Arg
 Pro
 Gln
 Asp
 Val
 Lys
 Phe
 Pro
 Gly
 Gly
 Gly
 Gly
 Pro
 Lys
 Phe
 Pro
 Gly
 Gly
 Gly
 Pro
 Lys
 Pro
 Gly
 Gly
 Pro
 Lys
 Pro
 Gly
 Pro
 Lys
 Ser
 Gly
 Pro
 Instance
 Fro
 Instance
 Instanc

35

30

Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 105 110 100 Arg His Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val 130 135 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 160 165 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 180 Ser Cys Ile Ser Val Pro Val Ser Ala 190 185

### (2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: HK10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

Met Ser Thr Leu Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Ile Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 2.0 25 Gln Ile Val Gly Gly Val Tyr Val Leu Pro Arg Arg Gly Pro 30 35 40 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 55 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 70 60 65 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro 80 75 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 95 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 100 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 125 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val

135

35

0

5

10

15

20

25

30

Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly Val Arg Ala Leu Glu Asp Gly Ile Asn Phe Ala Thr Gly Asn 155 160 165 165 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Phe 170 Ser Cys Leu Ile His Pro Ala Ala Ser

5 Ser Cys Leu IIe His Pro Ala Ala Ser

#### (2) INFORMATION FOR SEQ ID NO: 188:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

# (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: S52

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188

30 40 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser

Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
60 65 70
Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro
75 80

Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu
85 90 95
Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro
100 105 110
Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
115 120
Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val

130 135 140
Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly
145 150

Val Arg Ala Leu Glu Asp Gly Ile Asn Phe Ala Thr Gly Asn 155 160 165

Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Phe 170 175 180

Ser Cys Leu Val His Pro Ala Ala Ser

372577 1

35

30

0

10

	(2)	IN	IFORM	IATIC	N FC	R SE	EQ II	NO:	18	9:						
5		(	(i)		SEQUENCE CHARACTERISTICS: (A) LENGTH: 191 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: unknown (D) TOPOLOGY: unknown											
		(	(vi)		(A)	ORIGINAL SOURCE: (A) ORGANISM: homosapiens (C) INDIVIDUAL ISOLATE: S2										
10			(xi)		SEQU	SEQUENCE DESCRIPTION: SEQ ID NO: 189										
	Met	Ser	Thr	Leu	Pro 5	Lys	Pro	Gln	Arg	Lys 10	Thr	Lys	Arg	Asr		
		Ile	Arg	Arg		Gln 20	Asp	Ile	Lys		Pro 25	Gly	Gly	Gly		
		Ile 30	Val	Gly	Gly	Val	Tyr 35	Val	Leu	Pro	Arg	Arg 40	Gly	Pro		
15	Arg	Leu	Gly 45	Val	Arg	Ala	Thr	Arg 50	Lys	Thr	Ser	Glu	Arg 55	Ser		
	Gln	Pro	Arg	Gly 60	Arg	Arg	Gln	Pro	Ile 65	Pro	Lys	Ala	Arg	Arc 70		
		Glu	_		75					80						
20	85	Tyr	_			90					95					
		Pro 100	_	_		_	105					110				
		Arg	115		_			120					125			
		Cys		130					135					140		
25	_	Ala			145					150						
	155	Arg				160					165					
		Pro 170	_	_			175			Leu	Leu	Ala 180	Leu	Phe		
	Ser	Cys	Leu 185	Ile	His	Pro	Ala	Ala 190	Ser							

0

INFORMATION FOR SEQ ID NO: 190: (2)

> SEQUENCE CHARACTERISTICS:
> (A) LENGTH: 191 amino acids
> (B) TYPE: amino acid (i)

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: DK12

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

10 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser
45 50

Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
60 65 70

Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro
75 80

Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu

85 90 95 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 100 105 110

Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 125 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val

Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly
145
150

Val Arg Ala Leu Glu Asp Gly Ile Asn Phe Ala Thr Gly Asn 155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Phe 170 175 180

Ser Cys Leu Ile His Pro Ala Ala Ser 185 190

25

15

20

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids (B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

30

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: Z4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

0 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 5 10 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 5 50 Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 65 60 Pro Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 95 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 10 100 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Ile Val 135 130 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 15 Val Arg Ala Val Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 160 165 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 Ser Cys Leu Thr Val Pro Ala Ser Ala

# 20 (2) INFORMATION FOR SEQ ID NO: 192:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids

190

- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

# 25 (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: Z8
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 65 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro 80 75 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 105 100 5 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 130 135 140 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Ala Val Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 155 160 10 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 170 175 Ser Cys Leu Thr Val Pro Ala Ser Ala 185

(2) INFORMATION FOR SEQ ID NO: 193:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens
(C) INDIVIDUAL ISOLATE: Z1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 25 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly 20 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 40 Arg Leu Gly Val Arg Ala Ala Arg Lys Thr Ser Glu Arg Ser 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 70 30 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 100 105

35

15

Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Ala Val Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 165 160 5 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 170 Ser Cys Leu Thr Thr Pro Ala Ser Ala 185 190

# (2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: Z5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly 20 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Gln Ala Arg Arg 60 65 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 95 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Gln Asn Asp Pro 100 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 125 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 130 135 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arg Ala Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn

160

35

30

10

15

20

25

Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Phe 170 175 Ser Cys Leu Thr Thr Pro Ala Ser Ala 185

(2) INFORMATION FOR SEQ ID NO: 195:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly 15 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg 65 60 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro 20 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 95 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 100 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 125 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 25 135 130 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 150 145 Val Arq Ala Val Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 165 160 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu

(2) INFORMATION FOR SEQ ID NO: 196:

Ser Cys Leu Thr Val Pro Thr Ser Ala

(i) SEQUENCE CHARACTERISTICS:

175

35

30

170

c LENGTH: 191 amino acids (A) TYPE: amino acid (B) STRANDEDNESS: unknown (D) TOPOLOGY: unknown ORIGINAL SOURCE: (vi) (A) ORGANISM: homosapiens 5 (C) INDIVIDUAL ISOLATE: SEQUENCE DESCRIPTION: SEQ ID NO: (xi) 196: Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 5 10 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly 15 20 10 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 30 Arg Leu Gly Val Arg Thr Thr Arg Lys Thr Ser Glu Arg Ser 45 Gln Pro Arq Gly Arq Arq Gln Pro Ile Pro Lys Ala Arg Arg 60 65 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro 15 75 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 95 Ser Pro Arq Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 100 105 110 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 20 135 130 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Ala Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn 155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 180 170 Ser Cys Leu Thr Val Pro Ala Ser Ala 25 185

#### (2) INFORMATION FOR SEQ ID NO: 197:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

#### (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens(C) INDIVIDUAL ISOLATE: DK13

35

SEQUENCE DESCRIPTION: SEQ ID NO: 197: (xi) Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 5 10 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly 25 20 5 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 40 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 65 Leu Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 10 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 125 115 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val 15 135 130 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 Val Arg Leu Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu 175 Ser Cys Leu Thr Val Pro Ala Ser Ala 20

## (2) INFORMATION FOR SEQ ID NO: 198:

185

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

190

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: SA4

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

35

Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 65 60 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro 80 75 5 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 90 95 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro 100 105 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 115 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 10 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu 175 Ser Cys Leu Thr Val Pro Ala Ser Ala 15 185 190

#### (2) INFORMATION FOR SEQ ID NO: 199:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

### (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens(C) INDIVIDUAL ISOLATE: SA5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199: 25 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn ī 0 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 40 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 30 50 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 60 65 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro 75 80 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 90 85

35

Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro 105 110 100 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 120 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 165 160 155 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu 180 175 Ser Cys Leu Thr Val Pro Ala Ser Ala 190 185

10

15

0

5

# (2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

145

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: SA7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

150

20 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 30 35 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 25 50 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 60 65 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro 80 75 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro 30 110 105 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 140 130 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly

Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu 170 175 Ser Cys Leu Thr Val Pro Ala Ser Ala 185

5

10

- (2) INFORMATION FOR SEQ ID NO: 201:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 191 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown
  - (vi) ORIGINAL SOURCE:
    - (A) ORGANISM: homosapiens
      - (C) INDIVIDUAL ISOLATE: SA1
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:
- 15 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 5 10 Thr Asn Leu Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 4.0 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 50 20 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln
  - Fro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro
    75
    Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu
- 85 90 95 95

  Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro
  100 105 110

  Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
  - 115 120 125

    Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
    130 135 140

    Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly
    145 150
- 30 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu
  - 170 175 Ser Cys Leu Ile Ile Pro Ala Ser Ala 185 190

# (2) INFORMATION FOR SEQ ID NO: 202:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

# (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: SA3
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:
- 45 50 55

  15 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln
  60 65 70

  Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro

75 80

Leu Tyr Ala Asn Glu Gly Leu Glu Trp Ala Gly Trp Leu Leu 85 95

Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro

100 105 110

Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
115 120 125

Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
130 135 140

Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly
145
150

Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 25 155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu 170 175 180

Ser Cys Leu Thr Val Pro Ala Ser Ala 185 190

# 30 (2) INFORMATION FOR SEQ ID NO: 203:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

20

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

- (C) INDIVIDUAL ISOLATE: SA13
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:
- 5 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 1 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 15 Cln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 Gln Pro Arg Gly Arg Gln Pro Ile Pro Lys Ala Arg Gln Gln Pro Arg Gly Arg Gln Pro Ile Pro Lys Ala Arg Gln
  - Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro
    75

    Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu
    85

    90

    Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro
    100

    105
- 100 110 110 110 110 115 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 115 120 125 125 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val

Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 130 135 1440 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn

20 155 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu 170 175 Ser Cys Leu Thr Val Pro Thr Ser Ala

er Cys Leu Thr Val Pro Thr Ser Ala 185 190

## (2) INFORMATION FOR SEQ ID NO: 204:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
    (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: SA6
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Gln Arg Asn 1 5 10

35

Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly 20 25 15 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 40 Arg Met Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 5 65 Ser Ala Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro 75 8.0 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 95 90 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro 110 105 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 10 120 115 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 165 155 15 Leu Pro Gly Cys Ser Phe Ser Ile Phe Val Leu Ala Leu Leu 175 170 Ser Cys Leu Thr Val Pro Ala Ser Ala 185 190

## (2) INFORMATION FOR SEQ ID NO: 205:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 191 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: SA11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

35

25

30

Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro 80 75 Phe Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu 90 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro 110 105 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu 5 115 120 Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 135 130 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly 145 150 Arg Ala Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn 160 165 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu 10 170 175 Ser Cys Leu Thr Val Pro Ala Thr Ala 185 190

## (2) INFORMATION FOR SEQ ID NO: 206:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
      (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: HK2
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

Met Ser Thr Leu Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn 10 Thr Asn Arg Arg Pro Thr Asp Val Lys Phe Pro Gly Gly Gly 25 20 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro 35 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser 50 45 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln 60 6.5 70 Pro Gln Gly Arg His Trp Ala Gln Pro Gly Tyr Pro Trp Pro 75 80 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu 90 85 Ser Pro Arg Gly Ser Arg Pro His Trp Gly Pro Asn Asp Pro 105 110 100 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu

120

30

0

15

20

25

0	Thr	Cys	Gly	Phe 130	Ala	Asp	Leu	Met	Gly 135	Tyr	Ile	Pro	Val	Val 140	
	Gly	Ala	Pro	Leu	Gly 145	Gly	Val	Ala	Ala	Ala 150	Leu	Ala	His	Gly	
		Arg	Ala	Ile			Gly	Ile	Asn		Ala 165	Thr	Gly	Asn	
	155 Leu		Gly	Cys	Ser	160 Phe		Ile	Phe	Leu		Ala	Leu	Leu	
5	Ser	170 Cys	Leu 185	Thr	Thr	Pro	175 Ala	Ser 190	Ala			180			
	(2)	IN	FORM	ATIO	N FO	R SE	Q ID	NO:	207:						
10		(i	)		SEQUI (A) (B) (C) (D)	STR	CHAI GTH: E: : ANDE! OLOG	40 l nucle DNES	base eic : S:	pai acid sing	rs				
		(x	i)		SEQU	ENCE	DES	CRIP'	TION	: SE	Q ID	NO:	207:		
15	GCG	TCCG	GGT '	TCTG	GAAG.	AC G	GCGT	GAAC'	T AT	GCAA	CAGG				40
	(2)	IN	FORM	ATIO	N FO	R SE	Q ID	NO:	208:						
20		(i	)		SEQU (A) (B) (C) (D)	TYP STR	GTH:	40 nucl DNES	base eic	pai acid sing	rs				
		(x	i)		SEQU	ENCE	DES	CRIP	TION	: SE	Q ID	NO:	208:		
	AGG	CTTT	CAT	TGCA	GTTC	AA G	GCCG	TGCT	A TT	GATG	TGCC				40
25	(2)	IN	FORM	ATIC	N FO	R SE	Q ID	NO:	209:						
		(i	)		SEQU (A) (B) (C) (D)	TYP	CHA IGTH: E: ANDE OLOG	40 nucl DNES	base eic	pai acid sing	rs				
30		(x	i)		SEQU	ENCE	DES	CRIP	TION	: SE	Q ID	NO:	209:		
	AAG	ACGG	CGT	GAAC	TATG	CA A	CAGG	GAAC	C TT	CCTG	GTTC	;			40
	(2)	IN	FORM	ATIC	N FC	R SE	Q ID	NO:	210:						

372577\_1

0				
		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
5		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:210:	
	AGTTO	CAAGGC CGT	GCTATTG ATGTGCCAAC TGCCGTTGGT	40
	(2)	INFORMATIO	ON FOR SEQ ID NO:211:	
10		(i)	SEQUENCE CHARACTERISTICS: (A) LEMOTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:211:	
15	AAGA	CGGCGT GAA	TTCTGCA ACAGGGAACC TTCCTGGTTG	40
	(2)	INFORMATIO	ON FOR SEQ ID NO:212:	
20		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:212:	
	AGTT	CAAGGC CGT	GGAATTC ATGTGCCAAC TGCCGTTGGT	40
25	(2)	INFORMATI	ON FOR SEQ ID NO:213:	
		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 42 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
30		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:213:	
	ARCT	YCGACG TYA	CATCGAY CTGCTYGTYG GRAGYGCCAC CC	42
	(2)	INFORMATI	ON FOR SEQ ID NO:214:	
35	372577_1			

372577\_1

0				
		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 31 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
5		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:214:	
	RCAR	GCCRTC TTG	GAYATGA TCGCTGGWGC Y	31
	(2)	INFORMATI	ON FOR SEQ ID NO:215:	
10		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:215:	
15	CRAT	ACGACR YCA	YGTCGAY TTGCTCGTTG GGGCGGCTRY YT	42
	(2)	INFORMATI	ON FOR SEQ ID NO:216:	
20		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:216:	
	RCAA	GCTRTC RTG	GAYRTGG TRRCRGGRGC C	31
25	(2)	INFORMATI	ON FOR SEQ ID NO:217:	
		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
30		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:217:	
	TTGO	eggacke aca	ATYGACAT GGTYGTGATG TCCGCCACGC	4 (
	(2)	INFORMAT	ION FOR SEQ ID NO:218:	
35				

0		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 43 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
5		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:218:	
	GATGO	CGCGTT CCC	BAGGTCA TCWTAGACAT CRTYRGCGGR GCD	43
	(2)	INFORMATIO	ON FOR SEQ ID NO:219:	
10		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 54 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:219:	
15	AATGO ACAC	GCACCY TGCI	RCTGCTG GATACAAGTR ACACCTAATG TGGCTGTGAA	50 54
	(2)	INFORMATIO	ON FOR SEQ ID NO:220:	
20		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:220:	
	TGAR	CTAGYC CTY	SARGTYG TCTTCGGYGG Y	31
25	(2)	INFORMATI	ON FOR SEQ ID NO:221:	
30		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 54 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:221:	
	GCCAI TCAA		GATGTTG GGTGCCGGTT GCCCCCAATC TCGCCATAAG	5 ( 5 4

0				
Ü	(2)	INFORMATIO	ON FOR SEQ ID NO:222:	
5		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 46 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:222:	
	AAGG	GCCTGC GAG	CACACAT CGATATCATC GTGATGTCTG CTACGG	46
••	(2)	INFORMATIO	ON FOR SEQ ID NO:223:	
10		(i)	SEQUENCE CHARACTERISTICS: (A) LEMGTH: 45 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
15		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:223:	
13	TTGG'	rgcgca tcc	CGGAAGT CATCTTGGAT ATTGTTACAG GAGGT	45
	(2)	INFORMATI	ON FOR SEQ ID NO:224:	
20		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 40 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:224:	
25	AGTC.	AGGTAY GTC	GGAGCAA CCACCGCYTC GATACGCAGT	40
	(2)	INFORMATI	ON FOR SEQ ID NO:225:	
30		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 46 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:225:	
	AGCC	TTCACG TTC	AGACCKC GTCGCCATCA AACRGTCCAG ACCTGT	46

0				
	(2)	INFORMATI	ON FOR SEQ ID NO:226:	
5		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 75 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:226:	
			GTATGGT GGTRGCGCAC RTYCTGCGDY TGCCCCAGAC	50 75
10	(2)	INFORMATI	ON FOR SEQ ID NO:227:	
		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 39 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDMESS: single  (D) TOPOLOGY: linear	
15		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:227:	
	ACGC	CCGGTGA CGC	CCTACAGT GGCTGTCGCA CACCCGGGC	39
	(2)	TNEORMATI	ON FOR SEQ ID NO:228:	
20	(2)	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:228:	
25	ATGA	AGGGTCC CCA	ACAGCCTT TCTCGACATG GTTGCCGGAG GC	42
	(2)	INFORMAT	ION FOR SEQ ID NO:229:	
30		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:229:	
	CGCC	GCCCTAT CC	CAACGCAC CGTTAGAGTC CATGCGCAGG	40
35				

372577\_1

۰				
	(2)	INFORMATIO	ON FOR SEQ ID NO:230:	
5		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 49 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:230:	
	TCAG.	ATCTTA CGG	ATCCCCT CTATCCTAGG TGACTTGCTC ACCGGGGGT	49
10	(2)	INFORMATIO	ON FOR SEQ ID NO:231:	
		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 54 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
15		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:231:	
	CAGT CGGT		GGGTGGC CCTTACTCCC ACCGTGGCGG YGYCTTATAT	50 54
	(2)	INFORMATI	ON FOR SEQ ID NO:232:	
20		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 31 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
25		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:232:	
23	TAGC	ACTCTG GTR	GAYCTAC TCRCTGGAGG G	31
	(2)	INFORMATI	ON FOR SEQ ID NO:233:	
30		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 54 base pairs  (B) TYPE: nucleic acid  (C) STRANDENDESS: single  (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:233:	
35	AAGT	CTACAT GCT	GGGTGTC TCTCACCCCC ACCGTGGCTG CGCAACATCT	50
	372577_1			

372577\_1

۰				
	GAAT			54
	(2)	INFORMAT	ION FOR SEQ ID NO:234:	
5		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:234:	
10	AGGC	GCCATG GT	CGACCTGC TTGCAGGCGG C	31
10	(2)	INFORMAT	ION FOR SEQ ID NO:235:	
15		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 43 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
15		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:235:	
	TCAG	CCCCGA VY	YTCGGAGC GGTCACGGCT CCTCTTCGGA GGG	43
	(2)	INFORMAT	ION FOR SEQ ID NO:236:	
20		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 44 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
25		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:236:	
23	TGYT.	ACGGAT YC	CCCARGTG GTCATHGACA TCATWGCCGG GGSC	44
	(2)	INFORMAT	ION FOR SEQ ID NO:237:	
30		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:237:	
35	CATA	CCAAAT GC	TTCCACGC CCGCAACGGG ATTCCGCAGG	40

	(2)	INFO	RMATIO	N FOI	R SE	Q ID	NO:	238:						
5		(i)		SEQUI (A) (B) (C) (D)	TYPI STR	GTH: E: 1	37 l nucle ONES	ERIST pase eic a S: s linea	pain acid sing:	rs				
		(xi)		SEQUI	ENCE	DES	CRIP	rion:	SEQ	Q ID	NO:2	238:		
	TCTT	CTTGC	g ggcg	CCGC	AG T	GTT.	rgcT(	CATO	CCCT	3				37
10	(2)	INFO	RMATIO	N FOI	R SE	Q ID	NO:	239:						
		(i)		SEQUI (A) (B) (C) (D)	TYPI STR	GTH: E: 1 ANDEI	52 I nucle ONES	ERISToase eic a S: s linea	pain acid sing:	rs				
15		(xi)		SEQUI	ENCE	DES	CRIP	rion:	: SE(	QID	NO:2	239:		
	ATCT GC	FAGCAT	C TTGA	GGGT	AC C	rgag;	ATTT	G TGO	GAG'	rgtg	ATA	rttgo	GTG	50 52
	(2)	INFO	RMATIO	N FO	R SE	Q ID	NO:	240:						
20		(i)		SEQUI (A) (B) (C) (D)	TYPI STR	GTH: E: & ANDE	33 amino	amino	ac: id inkn	ids				
25		(xi)		SEQU!	ENCE	DES	CRIP'	rion:	SE	Q ID	NO:2	240:		
23	Trp	Ile G	ln Val	Thr	Pro	Asn	۷al	Ala		Lys	His	Arg	Gly	
	Leu	Thr H	is Asn		Arg	Xaa	His	Xaa		Xaa	Ile	Val	Met	
	Ala	Thr V	al	20					25					30
30	(2)	INFO	RMATIO	N FO	R SE	QID	NO:	241:						
		(i)		SEQUI (A) (B) (C) (D)	TYPI STR	GTH: E: : ANDE	33 a amino DNES	amino	ac: id inkno	ids				
35	372577	1												
	3,23//_	.*												

5

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

Trp Val Pro Val Ala Pro Asn Leu Ala Ile Ser Gln Pro Gly Ala 10 Leu Thr Lys Gly Leu Arg Ala His Ile Asp Ile Ile Val Met Ser 20 Ala Thr Val

- (2) INFORMATION FOR SEO ID NO:242:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 33 amino acids (B)
- TYPE: amino acid (C) STRANDEDNESS: unknown

  - (D) TOPOLOGY: unknown
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

Trp Ile Pro Val Xaa Pro Asn Val Ala Val Xaa Xaa Pro Gly Ala 10 15 Leu Thr Gln Gly Leu Arg Thr His Ile Asp Met Val Val Met Ser 2.0 25 Ala Thr Leu

- (2) INFORMATION FOR SEQ ID NO:243:
- (i) 20 SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:
- 25 Trp Thr Xaa Val Thr Pro Thr Val Ala Val Arg Tyr Val Gly Ala 10 Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val Gly Ala 20 25 Ala Thr Xaa
- (2) INFORMATION FOR SEO ID NO:244: 30
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 33 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

Trp Val Ala Leu Xaa Pro Thr Leu Ala Ala Arg Asn Xaa Xaa Xaa 5 10 10 15 Xaa Thr Xaa Xaa Ile Arg Xaa His Val Asp Leu Leu Val Gly Ala 20 25 30 Ala Xaa Phe

- (2) INFORMATION FOR SEQ ID NO:245:
  - (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 33 amino acids
    - (B) TYPE: amino acid (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

Trp Val Xaa Xaa Xaa Pro Thr Val Ala Thr Arg Asp Gly Lys Leu

10 15
Pro Xaa Xaa Gln Leu Arg Arg Xaa Ile Asp Leu Leu Val Gly Ser
20 25
Ala Thr Leu

- (2) INFORMATION FOR SEQ ID NO:246:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 33 amino acids
    - (B) TYPE: amino acid
      (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown
  - (xi) SEOUENCE DESCRIPTION: SEO ID NO:246:

Trp Thr Pro Val Thr Pro Thr Val Ala Val Ala His Pro Gly Ala
5 10 Pro Leu Glu Ser Phe Arg Arg His Val Asp Leu Met Val Gly Ala
20 25 Ala Thr Leu

- (2) INFORMATION FOR SEO ID NO:247:
- (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 33 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

35

30

5

10

15

Trp Val Ala Leu Thr Pro Thr Val Ala Xaa Xaa Tyr Ile Gly Ala
10 15
Pro Leu Xaa Ser Xaa Arg Arg His Val Asp Leu Met Val Gly Ala
20 25
Ala Thr Val

- 5 (2) INFORMATION FOR SEQ ID NO:248:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 33 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

Trp Val Ser Leu Thr Pro Thr Val Ala Ala Gln His Leu Asn Ala
5
Pro Leu Glu Ser Leu Arg Arg His Val Asp Leu Met Val Gly Gly
20
Ala Thr Leu

- (2) INFORMATION FOR SEQ ID NO:249:
  - (i) SEOUENCE CHARACTERISTICS:
    - (A) LENGTH: 33 amino acids
    - (B) TYPE: amino acid
- 20 (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
  - (xi) SEOUENCE DESCRIPTION: SEO ID NO:249:

Trp Val Pro Leu Thr Pro Thr Val Ala Ala Pro Tyr Pro Asn Ala
5 10 15
Pro Leu Glu Ser Met Arg Arg His Val Asp Leu Met Val Gly Ala
25 20 25 30

- Ala Thr Met
  - (2) INFORMATION FOR SEQ ID NO:250:
    - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 33 amino acids (B) TYPE: amino acid
      - (C) STRANDEDNESS: unknown
      - (D) TOPOLOGY: unknown
      - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

35

۰	Trp	Val	Xaa	Ile	Thr	Pro	Thr	Leu	Ser	Ala	Pro	Xaa	Xaa	Gly	Ala
	Val	Thr	Ala	Pro	Leu	Arg	Arg	Xaa	Val	Asp	Tyr	Leu	Ala	Gly	
	Δla	Δla	Len		20					23					50

(2) INFORMATION FOR SEQ ID NO:251:

5 (i) SEOUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

Trp His Ala Val Thr Pro Thr Leu Ala Ile Pro Asn Ala Ser Thr
5 10 15
Pro Ala Thr Gly Phe Arg Arg His Val Asp Leu Leu Ala Gly Ala
20 25 30
Ala Val Val

- 15 (2) INFORMATION FOR SEQ ID NO:252:
  - (i) SEOUENCE CHARACTERISTICS:
    - (A) LENGTH: 23 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro Glu Leu

Xaa Leu Xaa Val Phe Gly Gly
20

(2) INFORMATION FOR SEO ID NO:253:

(i)

- INFORMATION FOR SEQ ID NO.255.
  - SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

•

372577 1

35

25

30

٥															
	(2)	INI	FORM	ATIOI	I FOI	R SEÇ	Q ID	NO:	254:						
5		(i)	)		SEQUI (A) (B) (C) (D)	LENG TYPE STRA	FTH: E: & ANDEI	23 a amino ONES	ERIST amino aci S: u	aci d inkno	ids				
		(x:	i)	5	SEQUI	ENCE	DESC	CRIP'	rion:	SEÇ	Q ID	NO:2	254:		
	Thr	Xaa	Thr	Xaa	Ile 5	Leu	Ala	Tyr	Xaa	Met 10	Arg	Val	Pro	Glu	Val
10	Ile	Xaa	Asp	Ile		Xaa	Gly	Ala		10					15
	(2)	IN	FORM	OITA	v FOI	R SEC	Q ID	NO:	255:						
15		(i)	)		SEQUI (A) (B) (C) (D)	LENG TYPE STR	GTH: E: & ANDEI	23 a amino DNES	ERIST amino o aci S: u unkno	ac: id inkno	ids				
		(x:	i)	:	SEQUI	ENCE	DES	CRIP'	rion:	: SEQ	QID	NO:	255:		
	Ala	Val	Gly	Met	Val	Val	Ala	His	Xaa	Leu 10	Arg	Leu	Pro	Gln	Thr
20	Xaa	Phe	Asp	Ile	Xaa 20	Ala	Gly	Ala		10					13
	(2)	IN	FORM	ATIO	N FOI	R SE	Q ID	NO:	256:						
25		(i	)		SEQUI (A) (B) (C) (D)	LENG TYPI STR	GTH: E: & ANDE	23 amino	ERIST amino o ac: S: u	ac: id unkno	ids				
		(x	i)	:	SEQU	ENCE	DES	CRIP	rion	: SE	QID	NO:	256:		
	Thr	Xaa	Ala	Leu	Val 5	Xaa	Ser	Gln	Leu	Leu 10	Arg	Xaa	Pro	Gln	Ala 15
30	Xaa	Xaa	Asp	Xaa		Xaa	Gly	Ala							
	(2)	IN	FORM	ATIO	N FO	R SE	Q ID	NO:	257:						
35		(i	)						ERIS'						

372577\_1

- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown
- (xi) SEOUENCE DESCRIPTION: SEO ID NO:257:
- 5 Thr Xaa Ala Leu Val Xaa Ala Gln Leu Leu Arg Xaa Pro Gln Ala 5 10 15 Xaa Leu Asp Met Ile Ala Gly Ala 20
  - (2) INFORMATION FOR SEQ ID NO:258:
    - (i) SEQUENCE CHARACTERISTICS:
      - (A) LENGTH: 23 amino acids
      - (B) TYPE: amino acid
      - (C) STRANDEDNESS: unknown
      - (D) TOPOLOGY: unknown
    - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:
- Thr Thr Leu Leu Leu Ala Gln Ile Met Arg Val Pro Thr Ala
  5 10 15
  Phe Leu Asp Met Val Ala Gly Gly
  20
  - (2) INFORMATION FOR SEQ ID NO:259:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 23 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown

SEQUENCE DESCRIPTION: SEQ ID NO:259:

- (D) TOPOLOGY: unknown
- 25 Thr Thr Thr Leu Xaa Leu Ala Gln Val Met Arg Ile Pro Ser Thr
  5 10 15
  Leu Val Asp Leu Leu Xaa Gly Gly
  - 20

(xi)

30 (i) SEQUENCE CHARACTERISTICS:

INFORMATION FOR SEQ ID NO:260:

- (A) LENGTH: 23 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

35

10

20

(2)

Thr Ala Thr Leu Val Leu Ala Gln Leu Met Arg Ile Pro Gly Ala
5 10 15
Met Val Asp Leu Ala Gly Gly

- (2) INFORMATION FOR SEQ ID NO:261:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 23 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown
- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

Thr Ser Ala Leu Ile Met Ala Gln Ile Leu Arg Ile Pro Ser Ile
5 10 15
Leu Gly Asp Leu Thr Gly Gly
20

- 15 (2) INFORMATION FOR SEO ID NO:262:
  - (i) SEOUENCE CHARACTERISTICS:
    - (A) LENGTH: 23 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

Xaa Thr Ala Leu Xaa Met Ala Gln Xaa Leu Arg Ile Pro Gln Val 5 10 15 Val Ile Asp Ile Ile Ala Gly Xaa 20

- 25 (2) INFORMATION FOR SEQ ID NO:263:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 23 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS: unknown
    - (D) TOPOLOGY: unknown
- 30 (xi) SEQUENCE DESCRIPTION: SEO ID NO:263:

Thr Thr Leu Val Leu Ser Ser Ile Leu Arg Val Pro Glu Ile
5 10 15

Cys Ala Ser Val Ile Phe Gly Gly
20

- 243 -

## CLAIMS

- 1. A purified and isolated DNA having a sequence selected from the group consisting of SEQ ID NO:1 through SEO ID NO:51.
- A purified and isolated protein encoded by a gene whose sequence includes a sequence selected from the group consisting of SEO ID NO:52 through SEO ID NO:102.
- 3. A purified and isolated DNA having a sequence selected from the group consisting of SEQ ID NO: 103 through SEQ ID NO: 154.
  - A purified and isolated protein encoded by a gene sequence selected from the group consisting of SEQ ID NO: 155 through SEQ ID NO: 206.
    - 5. A purified and isolated protein having an amino acid sequence selected from the group consisting of SEQ ID NO:52 through SEQ ID NO:102 and SEQ ID NO:155 through SEQ ID NO:206.
    - 6. A method for the recombinant DNA-directed synthesis of a protein, said method comprising:

NO:206.

culturing a transformed or transfected host organism containing a DNA sequence capable of directing the host organism to produce said protein under conditions such that the protein is produced, said protein exhibiting substantial homology to a protein comprising the amino acid sequence selected from the group consisting of SEQ ID NO:52 through SEQ ID NO:102 or SEQ ID NO:155 through SEQ ID

30

0

5

15

20

15

- 7. The method of claim 6, wherein the host organism is transfected with a recombinant eukaryotic expression vector.
- 8. The method of claim 7, wherein the host  $\,$  organism is a eukaryotic cell.
- 9. A recombinant expression vector comprising a DNA sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:51 and SEQ ID NO:103 through SEQ ID NO:154.
  - 10. A host organism transformed or transfected with a recombinant expression vector according to claim 9.
  - 11. A method of detecting antibodies against  $\mbox{HCV},$  said method comprising:
    - (a) contacting a biological sample with at least one protein of claim 5 to form an immune complex with the antibodies; and
    - (b) detecting the presence of the immune complex.
- 12. The method of claim 11 wherein the biological sample is selected from the group consisting of serum, saliva or lymphocytes or other mononuclear cells.
  - 13. The method of claim 11, wherein the recombinant protein is bound to a solid support.
- 30 14. The method of claim 11, wherein the immune complex is detected using a labeled antibody.
- 15. A hepatitis C virus kit comprising: at least one protein comprising an amino acid sequence selected from the group consisting of: SEQ ID NO:52 through SEQ ID NO:102

and SEQ ID NO:155 through SEQ ID NO:206.

16. A composition comprising at least one recombinant protein of claim 5 and an excipient, diluent or carrier.

5 17. A composition comprising an expression vector capable of directing host organism synthesis of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO: 52 through SEO ID NO: 102 and SEQ ID NO: 155 through SEQ ID NO: 206.

10

A method of preventing hepatitis C infection, comprising administering the composition of claim 16 or 17 to a mammal in an effective amount to stimulate the production of protective antibody.

15

19. A vaccine for immunizing a mammal against hepatitis C infection, comprising at least one protein according to claim 5 in a pharmacologically acceptable carrier.

20

25

20. A vaccine for immunizing a mammal against hepatitis C infection, said vaccine comprising an expression vector capable of directing host organism synthesis of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:52 - SEQ ID NO:102 and SEQ ID NO:155 - SEO ID NO:206.

30

21. A method for detecting the presence of the hepatitis C virus via a reverse transcription-polymerase chain reaction, said method comprising amplifying an HCV reverse transcription product by polymerase chain reaction using universal primers.

35

22. The method of claim 21, wherein said universal primers are deduced from universally conserved 372577 1

nucleotide domains found in SEQ ID NO: 1 through SEQ ID NO: 51, in SEQ ID NO: 103 through SEQ ID NO: 154, or in consensus sequences shown in Figures 1A-H and 6A-K.

23. Substantially isolated and purified 5 universal primers, wherein said primers have nucleic acid sequences derived from universally conserved nucleotide domains found in SEQ ID NO:1 through SEQ ID NO:51, in SEQ ID NO:103 through SEQ ID NO:154 and in consensus sequences showing Figures 1A-H and 6A-K.

10

A diagnostic kit for use in detecting the presence of hepatitis C virus in a biological sample, said kit comprising at least two universal primers according to claim 22.

15

A diagnostic kit for use in detecting the presence of hepatitis C virus is a biological sample, said kit comprising at least one nucleic acid sequence selected from the group consisting of SEQ ID No:1-51 or SEO ID No:103-154

20

26. A method for determining the genotype of a hepatitis C virus, said method comprising:

25

amplifying reverse transcription products of RNA via polymerase chain reaction using genotype-specific amplification primers deduced from genotype-specific nucleotide domains found in SEQ ID NO:1 through SEQ ID NO:51, in SEQ ID NO:103 through SEO ID NO:154, or in consensus sequences shown in Figures 1A-H and 6A-K.

30

35

27. A method for determining the genotype of a hepatitis C virus, said method comprising:

5

10

15

20

25

- (a) amplifying RNA via reverse transcription-polymerase chain reaction to produce amplification products;
- (b) contacting said products with at least one sequence shown in SEQ ID NO:1 through SEQ ID NO:51 and SEQ ID NO:103 through SEQ ID NO:154; and
- (c) detecting complexes of said product which bind to said nucleic acid sequence.
- $$28.$\ A$  method for determining the genotype of a hepatitis C virus, said method comprising:
  - (a) amplifying RNA via reverse transcription-polymerase chain reaction to produce amplification products;
  - (b) contacting said products with at least one genotype-specific oligonucleotide; and
  - (c) detecting complexes of said products which bind to said oligonucleotide(s).
- 29. The method of claims 27 or 28, wherein said amplification of step (a) uses universal primers deduced from universally conserved nucleotide domains found in SEQ ID NO:1 through SEQ ID NO:51, in SEQ ID NO:103 through SEQ ID NO:154, or in consensus sequences shown in Figures 1A-H and 6A-K.
- 30. The method of claim 28, wherein said
  genotype-specific oligonucleotide of step (b) is a nucleic acid sequence deduced from genotype-specific nucleotide domains found in SEQ ID NO:1 through SEQ ID NO:51 and SEQ ID NO:103 through SEQ ID NO:154, or in consensus sequences shown in Figures 1A-H and 6A-K.

15

20

- 31. Substantially isolated and purified genotype-specific oligonucleotides, wherein said oligonucleotides have nucleic acid sequences deduced from genotype-specific nucleotide domains found in SEQ ID NO:1 through SEQ ID NO:51, in SEQ ID NO:103 through SEQ ID NO:154, or in consensus sequences shown in Figures 1A-H and 6A-K.
- 32. Substantially purified and isolated genotype-specific peptides having amino acid sequences

  10 deduced from a genotype-specific amino acid domains located in SEQ ID NO:52 through SEQ ID NO:102, in SEQ ID NO:155 through SEQ ID NO:206, or in consensus sequences shown in Figures 2A-H and 7A-K.
  - 33. A method of detecting antibodies specific for a single genotype of HCV, said method comprising:
    - (a) contacting a biological sample with at least one peptide of claim 32 to form an immune complex with the antibodies, and
    - (b) detecting the presence of the immune complex.
- 34. The method of claim 33, wherein the biological sample is selected from the group consisting of serum, saliva or lymphocytes or other mononuclear cells.
  - $$\,^{35}.$$  The method of claim 33, wherein said peptide is bound to a solid support.
  - 36. The method of claim 33, wherein the immune complex is detected using a labelled antibody or antigen.
- 37. A kit for use in detecting antibodies specific for a single genotype of HCV, said kit comprising: 372371

15

2.5

- at least one peptide selected from the genotype-specific peptides of claim 32.
  - 38. Substantially purified and isolated universal peptides having amino acid sequences deduced from universally conserved amino acid domains found in SEQ ID NO:52 through SEQ ID NO:102, in SEQ ID NO:155 through SEQ ID NO:206, or in consensus sequences shown in Figures 2A-H and 7A-K.
- 39. A method of detecting antibodies against all genotypes of HCV, said method comprising:
  - (a) contacting a biological sample with at least one peptide of claim 38 to form an immune complex with the antibodies, and
  - (b) detecting the presence of the immune complex.
- 40. The method of claim 39, wherein the biological sample is selected from the group consisting of serum, saliva or lymphocytes or other mononuclear cells.
  - 41. The method of claim 39, wherein said peptide is bound to a solid support.
  - 42. The method of claim 39, wherein the immune complex is detected using a labelled antibody or antigen.
- \$43.\$ A composition comprising at least one peptide of claim 32 and an excipient, diluent or carrier.
  - 44. A composition comprising at least one peptide of claim 38 and an excipient, diluent or carrier.
- \$45 . A method of preventing hepatitis C  $$^{372577}_{,1}$$

32. Substantially purified and isolated genotype-specific peptides having amino acid sequences deduced from a genotype-specific amino acid domains located in SEQ ID NO:52 through SEQ ID NO:102, in SEQ ID NO:155 through SEQ ID NO:206, or in consensus sequences shown in Figures 2A-H and 7A-K.

10

15

20

- 33. A method of detecting antibodies specific for a single genotype of HCV, said method comprising:
  - (a) contacting a biological sample with at least one peptide of claim 32 to form an immune complex with the antibodies, and
  - (b) detecting the presence of the immune complex.
- 34. The method of claim 33, wherein the biological sample is selected from the group consisting of serum, saliva or lymphocytes or other mononuclear cells.
  - 35. The method of claim 33, wherein said peptide is bound to a solid support.
  - 36. The method of claim 33, wherein the immune complex is detected using a labelled antibody or antigen.
- 37. A kit for use in detecting antibodies specific for a single genotype of HCV, said kit comprising: 372577\_1

- at least one peptide selected from the genotype-specific peptides of claim 32.
  - 38. Substantially purified and isolated universal peptides having amino acid sequences deduced from universally conserved amino acid domains found in SEQ ID NO:52 through SEQ ID NO:102, in SEQ ID NO:155 through SEQ ID NO:206, or in consensus sequences shown in Figures 2A-H and 7A-K.
- $_{
  m 10}$  39. A method of detecting antibodies against all genotypes of HCV, said method comprising:
  - (a) contacting a biological sample with at least one peptide of claim 38 to form an immune complex with the antibodies, and
  - (b) detecting the presence of the immune complex.
- 40. The method of claim 39, wherein the biological sample is selected from the group consisting of serum, saliva or lymphocytes or other mononuclear cells.
  - $\ensuremath{\mbox{41}}.$  The method of claim 39, wherein said peptide is bound to a solid support.
  - 42. The method of claim 39, wherein the immune complex is detected using a labelled antibody or antigen.
- \$43.\$ A composition comprising at least one \$30\$ peptide of claim 32 and an excipient, diluent or carrier.
  - 44. A composition comprising at least one peptide of claim 38 and an excipient, diluent or carrier.
- 35 45. A method of preventing hepatitis C  $_{372577\_1}$

- o infection, comprising administering the composition of claims 43 or 44 to a mammal in an effective amount to stimulate production of a protective antibody.
- 46. A vaccine for immunizing a mammal against 5 hepatitis C infection, comprising at least one peptide according to claims 32 or 38 in a pharmaceutically acceptable carrier.
- 47. A composition comprising at least one

  expression vector capable of directing host organism synthesis of a genotype-specific peptide having amino acid sequence deduced from a genotype-specific amino acid domain located in SEQ ID NO:52 SEQ ID NO:102, and SEQ ID NO:155 SEQ ID NO:206, or in consensus sequences shown in figures

  2A-H and 7A-K.
  - 48. A composition comprising at least one expression vector capable of directing host organism synthesis of a universal peptide having amino acid sequence deduced from universally conserved amino acid domains found in SEQ ID NO:52 SEQ ID NO:102, and SEQ ID NO:155 SEQ ID NO:206, or in consensus sequences shown in figures 2A-H and 7A-K.
- 49. A method of preventing hepatitis C infection, comprising administering the composition of claims 47 or 48 to a mammal in an effective amount to stimulate production of a protective antibody.
- 50. A vaccine for immunizing a mammal against hepatitis C infection, said vaccine comprising at least one expression vector capable of directing host organism synthesis of a geno-type specific peptide having amino acid sequence deduced from a geno type-specific amino acid domain located in SEQ ID NO:52 SEQ ID NO:102, and SEQ ID

- NO:155 SEQ ID NO:206, or in consensus sequences shown in figures 2A-H and 7A-K.
- 51. A vaccine for immunizing a mammal against hepatitis C infection, comprising at least one expression vector capable of directing host organism synthesis of a universal peptide having amino acid sequence deduced from universally conserved amino acid domain found in SEQ ID NO:52 SEQ ID NO:102, and SEQ ID NO:155 SEQ ID NO:206, or in consensus sequences shown in figures 2A-H and 7A-K.

52. Anti-HCV core antibodies having specific binding affinity for core protein of a single genotype of HCV.

- 53. Anti-HCV envelope 1 antibodies having specific binding affinity for envelope 1 protein of a single genotype of HCV.
- \$ 54. The antibodies of claims 52 or 53 wherein  $_{\rm 20}$   $\,$  said antibodies are monoclonal antibodies.
  - 55. A method of detecting core protein specific for a single genotype of HCV, said method comprising:
    - (a) contacting a biological sample with at least one antibody of claim 52 to form an immune complex with said core protein, and
      - (b) detecting the presence of the immune complex.
  - $\,$  56. A method of detecting E1 protein specific for a single genotype of HCV, said method comprising:
    - (a) contacting a biological sample with at least one antibody of claim 53 to form an immune complex with said E1 protein;

372577\_1

35

10

15

25

and

- (b) detecting the presence of the immune complex.
- 57. The methods of claims 55 or 56, wherein the biological sample is selected from the group consisting of serum, saliva lymphocytes or other mononuclear cells and liver.
- 58. The method of claims 55 or 56, wherein said 10 antibody is bound to a solid support.
  - 59. A method of detecting antibodies against all genotypes of HCV, said method comprising:
    - (a) contacting a biological sample with at least one universal peptide of claim 38 to form an immune complex with said antibodies; and
    - (b) detecting the presence of the immune complex.

20

15

0

25

## ABSTRACT

The nucleotide and deduced amino acid sequences of cDNAs encoding the envelope 1 genes and core genes of isolates of hepatitis C virus (HCV) are disclosed. The invention relates to the oligonucleotides, peptides and recombinant envelope 1 and core proteins derived from these sequences and their use in diagnostic methods and vaccines.

372577\_1

# FIGURE 1A

CEO ITO	NO. T1		
SEO ID	NO: Isolate S14	1	TACCAAGTGCGCAACTCCACGGGGCTTTACCATGTLACCAATGATTGCCCTAACTCGAGTA
_			
1	DK7	1	TACCAAGTGCGCAACTCCACGGGGCTTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
9	US11	1	TACCAAGTACGCAACTCCACGGGGCTTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
-			
4	DR4	1	CACCAAGTGCGCAACTCTACAGGGCTTTACCATGTCACCAATGATTGCCCTAATTCGAGTA
3	DR1	1	CACCAAGTGCGCAACTCTACAGGGCTTTACCATGTCACCAATGATTGCCCTAATTCGAGTA
			_
2	DK9	1	TACCAAGTACGCAACTCCCCGGGCCTCTACCATGTCACCAATGATTGCCCTAACTCGAGTA
6	S18	1	TACCAAGTACGCAACTCCACGGGCCTTTACCATGTCACCAATGACTGCCCTAACTCGAGCA
7	SW1	1	TACCAAGTACGCAACTCCLCGGGCCTTTACCATGTCACCAATGALTGCCCTAACTCGAGLA
1-8	consensus		tACCAAGT-CGCAACTCcaCgGGgCTtTACCATGTcACCAATGAtTGCCCTAACTCGAGtA
SEO ID	NO: Isolate		
5	S14	62	TEGTGTACGAGaCaGCEGATGCEATCCTaCACGCTCCGGGaTGTGTCCCTTGCGTTCGEGA
1	DK7	62	Tegrgtacgaggcggccgargccarccrgcacacrccggggrgtgrccctrgcgra
8	US11	62	TTGTGTACGAGGCGGCCGATGCCATCCTGCACACTCCGGGGTGTGTtCCTTGCGTTCGCGA
4	DR4	62	TTGTGTACGAGGCGGCGATGCCATCCTGCACACGCCGGGGTGTGTCCCTTGCGTTCGCGA
3	DR1	62	TTGTGTACGAGGCGGCCGATGCCATCCTGCACGCGCGGGGTGTGTCCCCTTGCGTTCGCGA
2	DK9	62	TTGTGTACGAGGCGGCCGATGCCATCCTGCALTCTCCaGGGTGTGTCCCTTGCGTTCGCGA
6	S18	62	TTGTGTACGAGACGGCCGATaCCATCCTACACTCTCCGGGGTGTGTCCCTTGCGTTCGCGA
7	SW1		
,	2#1	62	TTGTGTACGAGACGGCCGATgCCATtCTACACTCTCCaGGGTGTGTCCCTTGCGTTCGCGA
1 - 8	consensus		TtGTGTACGAGGCGGCCGATGCCATcCTgCAc-CtCCgGGGTGTGTcCCTTGCGTTCGcGA
SEO ID	NO: Isolate		
5	S14	123	GGGTAACAcCTCGAGGTGTTGGGTGGCGATGACCCCCACGGTGGCCACCAGGGAcGGCAAA
1	DK7	123	GGGTAACGECTCGAGGTGTTGGGTGGCGATGACCCCCACGGTGGCCACCAGGGAEGGCAAA
=			
8	US11	123	GGGTAACGCtTCGAGGTGTTGGGTGGCGATGACCCCCACGGTGGCCACCAGGGACGGCAAA
4	DR4	123	GGGTAACaCCTCGAGGTGTTGGGTGGCGGTGACCCCCACGGTGGCCACCAGGGACGGCAAA
3	DR1	123	GGGTAACGCCTCGAGGTGTTGGGTGGCGGTGACCCCCACGGTGGCCACCAGGGACGCAAL
2	DK9	123	GGGTAACGCCTCGA&ATGTTGGGTGGCGGTGGCCCCCACGGTGGCCACCAGGGACGGCAAG
_			
6	S18	123	GGGTAACGCCTCGAgATGTTGGGTGCCGGTGGCCCCCACAGTEGCCACCAGGGACGGCAAA
7	SW1	23	
1-8	consensus		GGgTaaCgcctCGAggTGTTGGGTGgCGgTGaCCCCCACgGTgGCCACcAGGGAcGGCAAa

FIGURE 1A

SEO ID 1	NO: <u>Isolate</u> S14	184 CTCCCCgCAaCGCAGCTTCGACGTtACATCGATCTGCTtGTCGGGAGCGC	CACCCTCTGTT
-			11111111111
1	DK7	184 CTCCCACAGCGCAGCTTCGACGTCACATCGATCTGCTcGTCGGGAGCGC	LACCCTCTGTT
8	US11	184 CTCCCCACAACGCAACTTCGACGTCACATCGATCTGCTTGTCGGGAGCGC	CACCCTCTGTT
4	DR4	184 CTCCCCACAACGCAGCTCCGACGTCACATCGACCTGCTTGTCGGGAGCGC	CACCCTCTGCT
3	DR1	184 CTCCCCACAACGCAGCTTCGACGTCACATCGACCTGCTTGTCGGGAGCGC	CACCCTCTGCT
2	DK9	184 CTCCCGCAACGCAGCTTCGACGTCACATCGATCTGCTTGTCGGGAGCGC	CACCCTCTGCT
6	S18	184 CTCCCGGAACGCAGCTTCGACGTCACATCGATCTGCTTGTtGGGAGCGC	11111111111
7	SW1		
•		CTCCC-CAACGCAGCT+CGACGT+ACATCGA+CTGCT+GT+GGGGAGGGC	
1-8	consensus	CICCCO-CARCGLAGGILCACAICAACCIGGICGGGAGGGG	CACCCICIGCI
SEO ID I	NO: Isolate		
5	S14	245 CGGCCCTCTACGTGGGGGACtTGTGCGGGTCTGTCTTTCTTGTCGGTCAG	
1	DK7	245 CGGCCCTCTACGTGGGGGACCTGTGCGGGTCTGTCTTTCTT	CTGTTTACCTT
8	S11	245 CGGCCTCTACGTGGGGGACCTGTGCGGGTCTGTCTTTCTT	
4	DR4	245 CGGCCCTCTACGTGGGGGACLTGTGCGGGTCTGTCTTCCTTGTCGGTCAA	CTGTTCACCTT
3	DR1	245 CGGCCTCTACGTGGGGGACCTGTGCGGGTCTGTCTTCCTTGTCGGTCAA	
2	DK9	245 CGGCCCTCTATGTGGGGGGACtTGTGCGGGTCTGTCTTCCTTGTCGGCCAA	11111111
6	S18	245 CGGCCCTCTATGTGGGGGGCCTGTGCGGGTCTGTCTTTCTT	
7	SW1	245 CGGCCCTCTAcGTGGGGGACLTGTGCGGGTCTGTCTTTCTcGTCAGLCAA	CTGTTCACgtT
1 - 8	consensus	CGGCCCTCTAcGTGGGGGAC-TGTGCGGGTCTGTCTTtCTtGTCgGtCAa	.CTGTTcACctT
SEO ID 1	NO: <u>Isolate</u> S14	306 CTCTCCCAGGCGCCtCTGGACGACGCAAGaCTGCAATTGTTCTATCTATC	C-CCCCATATA
5	514	306 CICICCAGGCGCCCCIGGACGACGCAAGACIGCAATIGTICTATCTATC	
1	DK7	306 CTCTCCCAGGCGCCACTGGACGACGCAAGGCTGCAATTGTTCTATCTA	
8	S11	306 CTCTCCCAGACGCCACTGGACGCCGCAGTGCAATTGTTCTATCTA	
4	DR4	306 CTCTCCCAGGCaCCACTGGACAACGCAAGACTGCAATTGTTCCATCTATC	
3	DR1	306 tTCTCCCAGGCGCCACTGGACAACGCAAGACTGCAATTGTTCTATCTA	
2	DK9	306 CTCCCCAGaCGCCACTGGACAACGCAAGACTGCAACTGTTCTATCTACC	CCGGCCATATE
6	S18	306 CTCCCCAGGCGCCACTGGACAACGCAAGACTGCAACTGTTCTATCTA	
			1111111111111
7	SW1	306 CTCCCCCAGGCGCCACTGGACAACGCAAGACTGLAACTGTTCTATCTALC	

FIGURE 1A

SEO ID NO:	<u>Isolate</u> S14	367	ACGGGTCAtCGCATGGCATGGGATATGATGATGACTGGTCCCCTACGACGGCacTGGTAG
-			
1	DK7	367	ACGGGTCACCGCATGGCGTGGGATATGATGATGATGACTGGTCCCCTACcACGGCGTTGGTAG
8	S11	367	ACGGGTCACCGCATGGCATGGGATATGATGATGAACTGGTCCCCTACGGCGGTGGTGG
4	DR4	367	
3	DR1	367	
2	DK9	367	ACGGGTCAtCGcATGGCGTGGGATATGATGATGATGATCCCCTACAGCAGCGCTGGTAA
6	S18	367	
7	SW1	367	ACGGGTCACCGCATGGCATGGGATATGATGATGATGGTCCCCCACAACAGCGCTGGTAg
1-8	consensus		${\tt ACGGGtCAcCGcATGGCaTGGGATATGATGATGAACTGGTCCCCtACgaC-GCgcTGGTag}$
	_		
SEO ID NO:	<u>Isolate</u> S14	428	TAGCTCAGCTGCTCCGGATCCCaCAAGCCATCTTGGALATGATCGCTGGTGCTCACTGGGG
-			
1	DK7	428	TAGCTCAGCTGCTCCGGATCCCGCAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGG
8	S11	428	TAGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGG
4	DR4	428	TAGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCCCACTGGGG
3	DR1	428	TGGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGAGCCCACTGGGG
2	DK9	428	TGGCgCAGCTGCTCAGGATCCCGCAgGCCATCTTGGACATGATCGCTGCTGCCCACTGGGG
6	S18	420	TAGCTCAGCTGCTCAGGGTCCCGCAAGCCGTCTTGGACATGATCGCTGGTGCCCACTGGGG
			111111111111111111111111111111111111111
7	SW1	428	TAGCTCAGCTGCTCAGGaTCCCGCAAGCCGTCTTGGACATGATCGCTGGTGCCCACTGGGG
1 - 8	consensus		${\tt TagCtCAGCTGCTCcGGaTCCC-CAaGCCaTCTTGGAcATGATCGCTGGtgCcCACTGGGGG}$
SEO ID NO:	<u> Isolate</u>		
5	S14	489	AGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGAAACTGGGCGAAGGTCCTAGTGGTG
1	DK7	489	AGTCCTgGCGGGCATAGCGTATTTtTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTG
8	S11	400	AGTCCTAGCGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTG
•	511		
4	DR4	489	AGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTG
3	DR1	489	AGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCGTGGTAGTG
2	DK9	489	AGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCGTGGTgGTa
6	S18	489	AGTCCTAGCGGGCATAGCGTATTTCTCCATGGGGGGAACTGGGCGAAGGTCCTGCTAGTG
7	SW1	489	AGTCCTAGCGGGCATAGCGTATTTCTCCATGGLGGGGGAACTGGGCGAAGGTCCTGaTAGTG
1-8	consensus		AGTCCTaGCGGGCATAGCGTATTTcTCCATGGtGGGGAACTGGGCGAAGGTCcTggTaGTg

## FIGURE 1A

SEO ID NO:	Isolate		
5	S14	550	CTGCTGCTATTcGCCGGCGTtGACGCG
1	DK7	550	CTGCTGCTATTTGCCGGCGTCGACGCG
8	US11	550	CTGCTGCTATTTGCCGGCGTCGACGCG
4	DR4	550	CTGTTGCTGTTTGCCGGCGTTGATGCG
3	DR1	550	CTGTTGCTGTTTGCCGGCGTTGATGCG
_			
2	DK9	550	CTGTTGCTGTTTaCCGGCGTCGATGCG
6	S18	550	CTGTTGCTGTTTqCCGGCGTCGATGCG
-			
7	SW1	550	CTGTTGCTGTTTtCCGGCGTCGATGCG
1-8	consensus		CTGtTGCTgTTtgCCGGCGTcGAtGCG

FIGURE 1B

SEO ID NO:	Isolate	
11	DK1	1 TATGAAGTGCGCAACGTGTCCGGGGTGTACCACGTCAC&AACGACTGCTCCAACTCAAGCA
24	T10	1 TATGAAGTGCGCAACGTGTCCGGGATGTACCALGTCACGAACGACTGCTCCAACTCAAGCA
10	D3	1 TATGAAGTGCGCAACGTGTCCGGGGTGTACCAAGTCACCAALGACTGTTCCAACTCGAGCA
9	D1	1 TATGAAGTGCGCAACGTGTCCGGGGTGTACCATGTCACGAACGA
14	HK5	1 TATGAAGTGCGCAACGTGTCCGGGGTATACCATGTCACGAACGA
	****	
15	HK8	1 TATGAAGTGCGCAACGTGTCCGGGATATACCATGTCACGAACGA
12	HK3	1 TATGAAGTGCGCAACGTGTCCGGGATATACCATGTCACGAACGA
12	nks	I INIGIA GEGARAGI GEGAGA TA LA CALCALA GALLA CALCALA GALLA G
23	Т3	1 TA-GAAGTGCGCAACGTGTCCGGGGTGTACLATGTCACGAACGACTGTTCCAACTCAAGCA
22	SW2	1 TATGAAGTGCGCAACGTGTCCGGGGTGTALCATGTCACGAACGACTGTTCCAACTCAAGCA
17	IND8	1 TATGAGGTGCGCAACGTGTCCGGGGTGTACCATGTCACGAACGA
16	IND5	1 TATGAAGTGCGCAACGTGTCCGGGGTGTACCATGTCACGAACGA
21	SA10	1 TATGAAGTGCGCAACGTGTCCGGGaTGTACCATGTCACGAACGACTGCTCCAACTCAAGCA
20	S45	1 TATGAAGTGCGCAACGTGTCCGGGGCGTACCATGTCACGAACGA
25	US6	1 TATGAAGTGCGCAACGTGTCCGGGATGTACCATGTCACGAACGA
25	036	1 TATGAAGTGCGCAACGTGTCCGGGATGTACCATGTCACGAACGA
13	HK4	1 cATGAAGTGCaCAACGTaTCCGGGATcTACCATGTCACGAACGACTGCTCCAACTCAAGTA
	1114	
18	P10	1 TATGAAGTGCGCAACGTGTCCGGGGTGTACCATGTCACGAACGA
19	S9	1 TATGAAGTGCGCAACGTaTCCGGGGGGTACCATGTCACGAACGACTGCTCCAACTCAAGTA
9-25	consensus	tAtGAaGTGCgCAACGTgTCCGGGgtgTAccAtGTCACgAAcGACTGcTCCAACTcaAGca

FIGURE 1B

SEO ID NO			
11	DK1	62	TCGTGTATGAGGCAGtGGACgTGATCATGCALACCCCAGGGTGCGTGCCCTGCGTTCGGGA
24	T10	62	TEGTGTETGAGGCAGCGGACETGATCATGCACCCCCGGGTGCGTGCCCTGCGTTCGGGA
10	D3	62	TCGTGTATGAGACAGCGGACATGATCATGCACACCCCCGGGTGCGTGC
9	D1	62	TEGTGTATGAGACAGCGGACATGATCATGCACACCCCCGGGTGCGTGC
14	HK5	62	TCGTGTACGAGACAACGGACATGATCATGCACACCCCTGGGTGCGTGC
15	HK8	62	TCGTGTATGAAACAGCGGACATGATLATGCATACCCCTGGATGCCATGCC
			<u> </u>
12	HK3	62	TCGTGTATGAGACAGCAGACATGATCATGCATACCCCTGGATGCGTGCCCTGCGTaCGGGA
			<u> </u>
23	<b>T</b> 3	62	TTGTGTATGAGACAGCGGACATGATCATGCACACCCCTGGGTGCCCTGCGTTCGGGA
			<u></u>
22	SW2	62	TTGTGTATGAGACAGCGGACATGATCATGCALACCCCCGGGTGCGTGCCCTGCGTTCGGGA
17			<u></u>
17	IND8	62	TTGTGTATGAGGCAGCGGACATGATCATGCACACCCCCGGGTGCGTGC
			<u> </u>
16	IND5	62	TTGTGTATGAGGCAGCGGACATGATCATGCACACCCCGGGTGCGTGC
			<u></u>
21	SA10	62	TTGTGTATGAGGCAGCGGACATGATCATGCACACCCCCGGGTGCGTGC
20	S45		<u> </u>
20	545	62	TTGTGTATGAGGCAGtGGACgTGATCcTGCACACCCCtGGGTGCGTGCCCTGCGTTCGGGA
25	US6		
25	036	62	TTGTGTATGAGGCAGCGGACATGATCATGCACACCCCGGGTGCGTGC
13	HK4		
13	HA4	62	TTGTGTATGAGGCAGCGGACATGATCATGCALACCCCCGGGTGCGTGCCCTGcGTcCGGGA
18	P10	-	
10	PIU	62	TTGTGTATGAGGCAGCGGACATGATAATGCACACCCCGGGTGCCCTGLGTTCGGGA
19	S9		
19	37	02	TTGTGTACGAGGCAGCGGGTGATCATGCALACCCCGGGTGLGTACCCTGCGTTCAGGA
9-25	consensus		TEGTGT=+GAGGCAGGACATGATGATGATGCA-ACGGCGGGGGGGGGG
2 - 23	COMPENSAR		TtGTGTatGAggCAgcgGACaTGATcaTGCAcACcCCcGGgTGcgTgCCCTGcGTtCgGGA

FIGURE 1B

SEO ID NO:	Isolate		
11	DK1	123	GaaCAACcaCTCCCGtTGCTGGGTAGCGCTCACCCCCACGCTCGCGGCCAGGAACgCCAGC
24	T10	122	
24	110	123	GGGCAACTCCTCCCGCTGCTGGGTAGCGCTCACCCCACGCTCGCGGCCAGGAACACCAGC
10	D3	123	GGACAACTCCTCTCGCTGCTGGGTAGCGCTCACCCCCACGCTCGCGGCTAGGAATAGCAGC
9	D1	123	GGACAACTCCTCTCGCTGCTGGGTAGCGCTCACCCCCACGCTCGCGGCTAGGAATGGCAaC
14	HK5	123	aAACAACTCCTCCCGTTGtTGGGTAGCGCTCGCCCCCACGCTCGCGGCCAGGAACGCCAGC
15	HK8	123	GAACAACTCCTCCCGTTGcTGGGTgGCGCTCACTCCCACGCTCGCGGCtAGGAAtGTCAGC
12	HK3	123	GAACAACTCCTCCCGCTGtTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACGTCAGC
23	Т3	123	GAGCAAtTCCTCCCGCTGCTGGGTAGCGCTtACTCCCACGCTCGCGGCCAGGAACGCCAGC
22	SW2	123	GGCCAACTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTaGCaGCCAGGAACaCCAGC
17	IND8	100	
17	IND8	123	GGGCAACTECTCTAGETGCTGGGTAGCGCTCACTCCCCACTCTCGCGGCLAGGAACGCCAGC
16	IND5	123	GGGCAACTCCTCTCGCTGCTGGGTAGCGCTCACTCCCCACTCTCGCGGCCAGGAACGCCAGC
21	SA10	123	GAACAACTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACTCCAGC
20	S45	122	GAACAACTCCTCCCGtTGCTGGGTgGCGCTCACTCCCACGCTCGCGGCCAGGAACTCCAGC
20	345	123	GAACAACTCCTCCCGtTGCTGGGTgGCGCTCACTCCCACGCTCGCGGCCAGGAACTCCAGC
25	US6	123	GAACAALTCCTCCCGcTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACGCLAGC
13	HK4	123	GAACAACTCCTCCCGtTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACGCCAGC
18	P10	123	GAACAACTCCTCCCGcTGCTGGTAGCGCTCACTCCCACaCTCGCGGCtAGGAAttCCAGC
19	S9	123	GggtAACTCCTCCCaaTGCTGGGTgGCGCTCACcCCCACgCTCGCGGCcAGGAAcgCtAcC
9 - 25	consensus		gaacAActcCTCccgcTGcTGGGTaGCGCTcaCtCCCACgCTcGCgGCcAGGAAcgccAgC

FIGURE 1B

SEO II	NO: Isolate		
11	DK1	184	aTCCCCACTACGACAATACGACGCCATGTCGATTTGCTCGTTGGGGGCGGCTGCTTTCTGCT
24			
24	T10	184	GTCCCCACTACGACGATACGACGCCATGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
10	. D3	194	GTCCCCACTACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTGTTTGCT
		101	
9	D1	184	GTCCCCACTACGGCGATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
14	HK5	184	GTCCCCACCACGCCAATACGACGCCACGTCGACTTGCTCGTTGGGGCGGCTGCTTTCTGCT
	*****		
15	HK8	184	GTCCCCACLACGACATACGACGCCACGTCGACTTGCTCGTTGGGGGCGGCTGCTTTCTGCT
12	HK3	184	GTCCCCACCACCACAACAACAACAACAACAACAACAACAAC
	111(3	101	
23	Т3	184	GTCCCCACTAGGACAATACGACGTCACGTCGACTTGCTCGTTGGGGCCCCCTTGCTTTCTG+T
22	SW2	184	GTCCCCACTACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
17	IND8	184	GTCCCCACCACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGTT
16	IND5	104	GTCtCCACCACGACAATACGACACCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGTT
10	11103	104	GICCCACCACGACACACACCACGTCGATTIGCTCGTTIGGGGCGGCTGCTTTCTGTT
21	SA10	184	GTCCCCACTACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCCGCTGCTTTTCTCCT
20	S45	184	GTCCCCACTACGACAATACGACGECACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
25			
25	US6	184	GTCCCCACTACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTaCTTTCTGCT
13	HK4	104	aTCCCCACTACGACAATACGACGCCATGTCGACTTGCTCGTTGGGGGGGCTGCTTTTCTGCT
	111/4	101	aTCCCCACTACGACACTACGACGCCATGTCGACTTGCTCGTTGGGGCGGCTGCTTTCTGCT
18	P10	184	GTCCCaACTACGGCAATACGACGCCATGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
19	S9	184	GTCCCCACCACGACAATACGACGECATGTCGATTTGCTCGTTGGGGCGGCTGETTTCTGCT
9-25			
2-25	consensus		gTCcCcACtAcGaCaATACGACgcCAcGTCGAtTTGCTCGTTGGGGCGGCTgctTTCTGcT

FIGURE 1B

SEO ID NO:	Isolate		
11	DK1	245	CCGCTATGTAcGTGGGGACCTCTGCGGATCcGTTTTCCTCGTCTCTCAGCTGTTCACCTT
24	T10	245	CCGCTATGTALGTGGGAGACCTCTGCGGATCTGTTTTCCTCGTCTCTCAGCTGTTCACCTT
10	D3	245	CCGCCATGTACGTGGGGGATCTLTGCGGATCTGTTTTCCTCGTCTCCCAGCTGTTCACCTT
9	D1	245	CCGCCATGTACGTGGGGGATCTGTGCGGATCTGTTTCCTCCCAGCTGTTCACCCT
14	HK5	245	CCGCTATGTACGTGGGGGATCTLTGCGGATCTGTTTTCCTCCTCTCCCAGCTGTTCACCTT
15	HK8	245	CCGCTATGTACGTGGGGGATCTCTGCGGGATCTGTTTTCCTCGTCTCCCAGCTGTTCACCTT
12	HK3	245	CCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTCCTLGTCTCCCAGCTGTTCACCTT
23	T3	245	CCGCTATGTACGTGGGGGATCTCTCCGGGATCTCTCCCCGGCTGTTCACTTT
22	SW2	245	CCGLTATGTACGTGGGGGATCTCTGCGGGTTTTTCCTCGTCTCCCAGCTGTTCACTTT
17	IND8	245	CCGCTATGTACGTGGGGGATCTCTGCGGGATCTGTTTTCCTLGTCTCCCAGCTGTTCACCTT
16	IND5	245	CCGCTATGTACGTGGGGGATCTATGCGGATCTGTTTTCCTcGTCTCCCAGCTGTTCACCTT
21	SA10	245	CCGCcATGTACGTGGGGGACCTCTGCGGATCTGTTTTCCTTGTCTCCCAGCTGTTCACCTT
20	S45	245	CCGCTATGTACGTGGGGGALCTCTGCGGGATCTGTTTCCTTGTLTCCCAGCTGTTCACCTT
25	US6	245	CCGCTATGTACGTGGGGGACCTCTGCGGgTCcGTTTTCCTCATCTCCCAGCTGTTCACCTT
13	HK4	245	CCGCcATGTACGTGGGAGATCTCTGCGGATCTGTCTCCTCGTCTCCCAGcTGTTCACCTT
18	P10	245	CCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTCTCCTCGTCTCCCAGCTGTTCACCTT
19	S9	245	CCGCTATGTACGTGGGGGACCTGTGCGGATCTGTTLTCCTCATCTCCCAGCTGTTCACCAT
9 - 25	consensus		CCGctATGTAcGTGGGGGAtCTcTGCGGaTCtGTttTCCTcgTcTCcCAGcTGTTCACctT

FIGURE 1B

SEQ ID NO:	Isolate		
11	DK1	6 tTCaCCTCGCCGGCATGAGACagcaCAGGACT	GCAACTGCTCAATCTATCCCGGCCAcgTt
24	TIO	6 CTCGCCTCGCCGGCATGAGACttTgCAGGACT	CGD D CTC CTC D TO CTC
24	110		GCAACIGCICAAICIAICCCGGCCAECIG
10	D3	6 CTCGCCTCGCCGGCATGAGACAGTACAGGAAT	GTAACTGCTCAATCTATCCCGGCCACGTG
9	D1	6 CTCGCCTCGCCGGCATGAGACGGTACAGGAGT	GTAALTGCTCAATCTATCCCGGCCACGTG
		<u> </u>	<u>                                     </u>
14	HK5	6 CTCGCCTCGCCGACACGAGACGGTACAGGACT	GCAACTGCTCAATCTATCCCGGCCACGTA
15	HK8	6 tTCGCCTCGCCGACACGAGACGGTACAGGACT	
15	nko		
12	HK3	6 CTCGCCTCGCCGACACGAGACAGTACAGGACT	GCAACTGCTCACTCTATCCCGGCCACGTA
23	T3	6 CTCGCCTCGCCGGCALGAGACAGTACAGGACT	GCAACTGCTCAATCTATCCCGGCCACGTA
		_ 11 11111111111 11111111111111111	1   1   1   1   1   1   1   1   1   1
22	SW2	6 LTCACCTCGCCGGCACGAGACAGTACAGGACT	GCAACTGETCCATCTATCCCGGCCACGTA
17	IND8	6 CTCACCGCGCCGGCATGAGACAGTACAGGACT	CCD DATE CARCAL MARKET DE COCCOCA COMP
	11120	U CICACCOCCOCCATGAGACAGIACAGGACI	!!!!!!!!!!!!!!!!!!!!!!!!!!!!
16	IND5	6 CTCACCGCGCCGGCATGAGACAGTACAGGACT	GCAATTGCTCCATCTATCCCGGCCACGTA
21	SA10	6 CTCGCCTCGCCGGLATGAGACAGTACAGGACT	GCAATTGCTCAATCTATCCCGGCCGCGTA
20	S45	6 CTCGCCTCGTCGGCATGAGACAGTACAGGACT	GCAACTGTTCAATCTATCCCGGCCACGTA
25	US6	6 CTCGCCTCGTCaGCATGAGACAGTACAGGACT	
	000		
13	HK4	6 CTCGCCTCGCCGGCATGAGACGGTACAGGACT	GCAATTGCTCAATCTATCCCGGCCACGTA
18	P10	6 CTCaCCTCGCCGGCATtgGACAGTACAGGACT	GCAATTGETCAATCTATCCEGGCCACGTA
19	S9		
13	59	6 CTCgCCcCGtCGGCATgaGACAGTACAGaACT	GCAATTGCTCAATCTATCCCGGaCACGTg
9-25	consensus	cTCgCCtCGcCggcAtgaGACagtaCAGgAcT	GCAACTGCTCaaTCTATCCcGGcCacqTa

FIGURE 1B

SEO I	D NO: Isolate		
11	DK1		TCAGGTCACCGCATGGCTTGGGALATGATGATGACTGGTCaCCTACAACAGCcCTAGTGC
24	T10	367	TCAGGTCACCGCATGGCTTGGGACATGATGATGAACTGGTCGCCTACAACAGCtCTAGTGG
10	D3	367	ACAGGTCACCGCATGGCTTGGGATATGATGATGACTGGTCGCCTACAGCACCCTAGTGG
9	D1	267	ACAGGTCACCGLATGGCTTGGGATATGATGATGAACTGGTCACCTACAACAGCCC+TACTGG
,	<b>D1</b>	307	ACASSICACCSCAIGGCIIGGGAIAIGAIGAIGAACIGGICACCIACAACAGCCCIAGIGG
14	HK5	367	ACAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAACAGCCCTAGTGG
		-	
15	HK8	367	TCAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCGCCCACAACAGCCCTAGTGG
12	HK3	367	TCAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCcCCLACAGCAGCCCTAGTGG
23	<b>T</b> 3	367	aCAGGTCACCGtATGGCTTGGGATATGATGATGAACTGGTCgCCcACAaCgGCaCTAGTGG
22	SW2		
22	Sw2	367	TCAGGTCACCGCATGGCTTGGGACATGATGATGACTGGTCACCTACAGCAGCCCTgGTGG
17	INDS	367	TCAGGTCACCGCATGGCTTGGGATATGATGATGACTGGTCACCTACAGCGGCCCTAGTGG
1,	11100	307	
16	IND5	367	TCAGGTCACCGCATGGCCTGGGATATGATGATGAACTGGTCACCTACAGCAGCCCTAGTGG
21	SA10	367	ACAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAACAGCtCTAGTaG
20	S45	367	ACAGGTCACCGCATGGCTTGGGATATGATGATGATCTGGTCGCCTACAGCAGCCLTAGTGG
25	US6	367	TCAGGTCACCGCATGGCTTGGGATATGATGATGAALTGGTCACCTACAGCAGCCCTAGTGG
13	HK4	367	TCAGGTCACCGCATGGCTTGGGATATGATGATGATCTCGCTCACCTACAGCAGCCCTACTTCC
13	nr.4	367	TCAGGTCACCGCATGGCTTGGGATATGATGATGACTGGTCACCTACAGCAGCCCTAGTGG
18	P10	367	TCAGGTCACCGCATGGCTTGGGATATGATGATGATGTGGTCGCCCACAGCCCCTAGTGC
		50.	IIIIII IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
19	S9	367	aCAGGTCAtCGCATGGCcTGGGATATGATGATGAACTGGTCGCCtACAACAGCCCTAGTGG
9-25	consensus		tCAGGTCAcCGcATGGCtTGGGAtATGATGATGAAcTGGTCaCCtACAgCaGCccTaGTgg

FIGURE 1B

SE	O ID NO:	Isolate		
	11	DK1	428	${\tt TATCGCAGTTACTCCGAATCCCACAAGCTGTCgTGGACATGGTGgCgGGGGCCCACTGGGG}$
	24	T10		
	24	110	428	TgTCGCAGTTACTCCGGATCCCACAAGCTGTCaTGGACATGGTGaCaGGGGCCCACTGGGG
	10	D3	428	TATCGCAGTTACTCCGGATCCCACAAGCTGTCgTGGACATGGTGGCGGGGGCCCACTGGGG
	_			
	9	D1	428	TATCGCAGTTACTCCGGATCCCACAAGCTGTCaTGGACATGGTGGCGGGGGCCCACTGGGG
	14	HK5	428	TGTCGCAGTTACTCCGGATCCCGCAAGCTGTCGTGGACATGGTaGCGCCCCCCAACTGGCCC
	15	HK8	428	TGTCGCAGTTACTCCGGATCCCGCAAGCTaTCGTGGACATGGTGGCGGGGGCCCACTGGGG
	12	нк3	420	TGTCGCAaTTACTCCGGATCCCGCAAGCTGTCGTGGACATCGTGGGGGGCCCACTGGGG
		nks	428	TGTCGCAATTACTCCGGATCCCGCAAGCTGTCGTGGACATGGTGGCGGGGCCCACTGGGG
:	23	Т3	428	TGTCGCAGTTgCTCCGGATCCCACAAGCTGTCGTGGACATGGTGGCGGGGCCCACTGGGG
-	22	SW2	428	${\tt TATCGCAGTTACTCCGGATCCCACAAGCTGTCGTGGACATGGTAGCGGGGGCCCACTGGGG}$
	.7	IND8	428	TATCGCAGTTGCTCCGGATCCCACAAGCTGTCGTGGATATGGTGGCGGGGCCCACTGGGG
1	.6	IND5	428	TATCGCAGTTGCTCCGGATCCCACAAGCTGTCGTGGATATGGTGGCGGGGGCCCACTGGGG
-	1	SA10	428	TRITCGCNGTTN CTCCCCCN CRARCE TO COMPANY COMPAN
_	-	57120		TATCGCAGTTACTCCGGATCCCACAAGCTaTCGTGGACATGGTGGCGGGGCCCACTGGGG
2	0	S45	428	TATCGCAGTTACTCCGGATCCCACAAGCTGTCGTGGACATGGTGGCGGGGGCCCACTGGGG
	:5	US6	428	
•		036	428	TATCGCAGTTACTCCGGATCCCACAAGCTGTCATGGACATGGTGGCGGGGGCCCACTGGGG
1	.3	HK4	428	TATCGCAGTTACTCCGacTCCCACAAGCTGTCATGGACATGGTGGCGGGGGCCCACTGGGG
_	_			
	.8	P10	428	TGTCGCAGCTACTCCGGATCCCACAAGCTATCLTGGATGTGGTGGCGGGGGCCCACTGGGG
1	.9	S9	428	TaTCGCAGCTACTCCGGATCCCACAAGCTgTCaTGGATaTGGTGGCGGGGGCCCACTGGGG
9 - 2	5 с	onsensus		TaTCGCAgtTaCTCCGgaTCCCaCAAGCTgTCgTGGAcaTGGTggCgGGGGCCCACTGGGG

FIGURE 1B

SEO I	NO: Isolate		
11	DK1	489	AGTCCTGGCGGGCCTcGCCTACTACTCCATGGCGGGGAACTGGGCcAAGGTTTTAATTGTG
24	T10	489	AGTCCTGGCGGGCCTEGCCTACTATTCCATGGCGGGAACTGGGCTAAGGTTTTAATTGTG
10	D3	489	GGTCCTGGCGGGCCTCGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTTTTGATTGTG
9	D1	489	GGTCTGGCGGCCTCGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTTTTGATTGTG
14	HK5	489	GGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGAAACTGGGCTAAGGTTTTGATTGTG
15	HK8	489	AGTCCTAGCGGGCCTTGCCTACTATTCCATGGTGGGCAACTGGGCTAAGGTTTTGATTGTG
12	HK3	489	AGTCCTAGCGGGCCTTGCCTACTATTCCATGGTGGGAAACTGGGCTAAGGTTTTGATTGTG
23	T3	489	AGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTTTTGATTGTG
22	SW2	489	AGTCCTGGCGGCCTTGCaTACTATTCCATGGTGGGGAACTGGGCTAAGGTTTTGATTGTG
17	IND8	489	AATCCTGGCGGCCTTGCCTACTATTCCATGGTAGGGAACTGGGCTAAGGTTTTGATTGTG
16	IND5	489	AATCCTGGCGGGCCTTGCCTACTATTCCATGGTAGGGAACTGGGCTAAGGTTTTGATTGTG
21	SA10	489	AGTCCTaGCGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTTTTGATTGTE
20	S45	489	AGTCCTGGCGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTTCTGATTGTG
25	US6	489	AGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTTCTGATTGTG
13	HK4	489	AGTCCTaGCGGGCCTTGCLTACTATTCCATGGTGGGGAACTGGGCCAAGGTTTTGATTGTG
18	P10	489	AGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTCTTGATTGTG
19	S9	489	AGTCCTGGCGGGCCTcGCCTACTATTCCATGGTGGGGAACTGGGCTAAGGTLTTGATTGTG
9-25	consensus		agTCCTgGCGGGCCTtGCcTACTAtTCCATGGtgGGGAACTGGGCtAAGGTttTgATTGTg

FIGURE 1B

SEO ID NO	: Isolate		
11	DK1	550	tTGCTACTCTTTGCCGGCGTTGATGGG
24	T10	550	ATGCTACTCTTTGCCGGCGTTGATGGG
10	. рз		
10	. Дз	550	ATGCTACTCTTTGCTGGCGTcGACGGC
9	D1	550	ATGCTACTCTTTGCTGGCGTTGACGGC
14	HK5	550	ATGCTACTETTTGCCGGCGTTGATGGG
15	HK8	550	ATGCTACTGTTTGCCGGCGTTGATGGG
		550	ATGCTACTGTTTGCCGGCGTTGATGGG
12	HK3	550	ATGCTACTTTTTGCCGGCGTTGATGGG
23	T3	550	CTGCTACTCTTTGCCGGCGTTGATGGG
22	SW2	550	777777777777777777777777777777777777777
17	INDS	550	ATGCTACTCTTTGCCGGCGTTGACGGG
16	IND5	550	ATGCTACTCTTTGCCGGCGTTGACGGG
21	SA10	550	ATGCTACTCTTTGCCGGCGTTGACGGG
20	S45	550	ATGCTACTCTTTGCCGGCGTTGACGGG
25	US6	550	tTGCTACTCTTTGCCGGCGTTGACGGG
13	HK4	550	ATGCTACTCTTTGCCGGCGTTGACGGG
18	P10	550	ATGCTACTCTTTGCCGGCGTTGACGGa
19	S9	550	ATGCTACTTTTTGCtGGtGTTGACGGg
9-25	consensus		aTGCTACTcTTTGCcGGcGTtGAcGGq

# FIGURE 1C

SEO ID NO	: Isolate T2	1 GCCCAAGTGAGGAACACCAGCCGCGGCTACATGGTGACCAACGACTGTTCCAATGAGAGCA
27	T4	
28	<b>T</b> 9	1 GCCgAAGTGAAGAACACCAGTACCAGCTACATGGTGACaAATGACTGTTCCAACGACAGCA
29	US10	1 GtCcAAGTGAA&AACACCAGTACCAGCTAtATGGTGACcAATGACTGcTCCAACGACAGCA
26-29	consensus	GcccAAGTGAagAACACCAgtacCaGcTAcATGGTGACcAA-GACTGtTCcAA-GACAGCA
SEO ID NO 26	: <u>Isolate</u> T2	62 TCACcTGGCAGCTCCAAGCCGGGGTtCTCCACGTCCCCGGGTGTATCCCGTGtGAGAggct
27	T4	62 TCACLTGGCAGCTCCAGGCCGCGTCCTCCACGTCCCCGGGTGTCTCCCGTGCGAGAAAac
28	Т9	62 TCACCTGGCAACTCCAGGCCGCGGTCCTCCACGTCCCCGGGTGCGTCCCGTGCGAGAGAGT
29	US10	62 TCACLTGGCAACTLGAGGCLGCGGTCCTCCACGTLCCCGGGTGLGTCCCGTGCGAGAAAGT
26-29	consensus	TCAC-TGGCA-CTccAgGCcGCGGTcCTCCACGTcCCCGGGTGtgTCCCGTGcGAGA-agt
SEO ID NO 26	<u>Isolate</u> T2	123 GGGAAATACATCCCGATGCTGGATACCGGTCACACCAAACGTGGCCGTGCGGCAGCCCGGC
27	T4	123 GGGAAATACATCtCGGTGCTGGATACCGGTtTCACCAAACGTGGCCGTGCGGCAGCCCGGC
28	<b>T</b> 9	123 tGGAAACGCGTCCTGGATACCGGTCTCGCCAAACGTAGCtGTGCAGCGGCCTGGC
29	US10	
26-29	consensus	$\tt gGGAAAtaCaTCtCGgTGCTGGATACCGGTctCaCCAAAcGTgGCcGTGC-GC-GGCGGCCGGCGGCGGGGGGGGGGGGGGGG$
SEO ID NO	Isolate T2	184 GCtCTtACGCAGGGCTTGCGGACGCACATcGACATGGTTGTGATGTCCGCCACGCTCTGCT
27	T4	184 GCCCICACGCAGGCTTGCGGACGCACATEGACATGGTTGTGATGTCCGCCACGCTCTGCT
28	T9	184 GCCCTCACGCAGGCTTGCGGACGCACATCGACATCGTTGTGTGTCTCCCCCCCC
29	US10	184 GCCCTCACGCAGGGCTTGCGGACLCACATCGACATGGTCGTGATGTCCGCCACGCTCTGCT
26-29	consensus	GC-CT-ACGCAGGCTTGCGGACGCACAT-GGACATGGT+GTGATGTCCGCCACGCTCTGCT
SEO ID NO:	Isolate T2	245 CTCCaCTaThCCTCCCCACCTCTCCCCACCTCTCCCCCCCCCC
27	T4	245 CTGCCCTCTACGTGGGGGACCTCTGGGGGGGGGTGATGCTCGCAGCCCAGATGTTCATEGT
28	Т9	
29	US10	245 CCGCTCTACGTGGGGGACTCTGCGGCGGGTAATGCTCGCGGCtCAGAATGTTCATTAT
	0310	245 CCGCTCTtTACGTGGGGACtTCTGCGGGGGGGTGATGCTCGCAGCCCAAATGTTCATTgT
26-29	consensus	C-GCtCT-TACGTGGGGGAccTCTGCGGcGGGGTGATGCTCGCaGCcCAgATGTTCATtgT

## FIGURE 1C

SEQ ID NO	: <u>Isolate</u> T2	206 00000000-2002000000000000000000000
27	T4	306 CTCGCCGGACGCCACTGGTTTGTGCAAGAATGCAATTGCTC-ATCTACCCGGGLACCATC
28	T9	306 CTCGCCGCAACAtCACTGGTTTGTGCAAGACTGCAATTGCTCTATCTACCCTGGCACCATC
29	US10	306 CTCGCCGCAgCACCACTGGTTTGTGCAGGAATGCAACTGCTCCATtTACCCTGGTACCATC
		306 CTCGCCGCgcCACCACTeGTTTGTGCAGGAATGCAACTGCTCCATeTACCCCGGTACCATC
26-29	consensus	CTCGCCGC-aCacCACTgGTTTGTGCA-GAATGCAA-TGCTCCATCTACCC-GGLACCATC
SEO ID NO	: Isolate	
26	T2	367 ACTGGACACCGTATGGCATGGGACATGATGAACTGGTCGCCCACAGCCACCATGATCC
27	T4	367 ACTGGACACCGTATGGCATGGGALATGATGATGAACTGGTCGCCCACGGCCACCATGATCC
28	Т9	367 ACTGGACACCGTATGGCATGGGACATGATGATGATGATCGTCGCCCACaaCCACCATGATCt
29	US10	367 ACCGGGCACCGTATGGCATGGGACATGATGATGATGATCGTCGCCCCACggCCACttTGATCC
26-29	consensus	ACtGGaCACCGTATGGCATGGGACATGATGATGAACTGGTCGCCCAC-gCCACcaTGATCc
SEO ID NO 26	<u>Isolate</u> T2	428 TGGCGTACGCGATGCGCGTTCCCGAGGTCATCATAGACATCATCGGCGGGCCCCACTGGGG
27	T4	428 TGGCGTACGCGATGCGCGTTCCCGAGGTCATCLTAGACATCGTLAGCGGGGCACACTGGGG
28	Т9	428 TGGCGTACGCGATGCGCGTTCCCGAGGTCATCATGAGCATCATCAGCGGAGCTCACTGGGG
29	US10	428 TGGCGTACGEGATGCGCGTTCCCGAGGTCATCATAGACATCATLAGCGGGGCgCALTGGGG
26-29	consensus	TGGCGTACGcGATGCGCGTTCCCGAGGTCATCaTAGACATCaT-aGCGGgGCtCAcTGGGG
000 TD 110		
SEO ID NO:	Isolate T2	489 CGTCATGTTEGGCTTGGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAGGTCATTGTCATC
27	T4	489 CGTCATGTTCGGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCCTTCTATATGCAGGAGCGTGGGCGAAAGTCCTTCTATATGCAGGAGCGTGGGCGAAAGTCCTTCTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTATATGCAGGAGGGGTGGGCGAAAAGTCCTTCTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTATATGCAGGAGGGTGGGGCGAAAAGTCCTTCTTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTTATATGCAGGAGGGTGGGGCGAAAAGTCCTTCTTATATGCAGGAGGGTGGGCGAAAAGTCCTTCTTATATGCAGGAGGGTGGGGCGAAAAGTCCTTCTTATATGCAGGAGGGAG
28	T9	489 CGTCATGTTCGGCCTAGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAGGTCGTTGTTATC
29	US10	489 CGTCtTGTTCGGCtTAGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTCTCATC
26-29	consensus	CGTCaTGTTcGGCtT-GCCTACTTCTCTATGCAGGGAGCGTGGGCGAA-GTCgTTGTCATC
		•
SEO ID NO:	<u>Isolate</u> T2	550 CTCtTGCTGGCtGCTGGGGTGGACGCG
27	Т4	
28	Т9	
29	US10	
26-29	consensus	CTt - TGCTggCcGCTGGgGTGGACGCG

# FIGURE 1D

SEC ID NO 33 30 32 31 30-33	DK8 SW3 DK11 consensus	1	GTGGAAGTLAGAAACACCAGTTLLAGCTACTACGCCACCAATGATTGCTCGAACAACACCA
SEO ID NO 33 30 32 31	T8 DK8 SW3 DK11	62 62	TCACCTGGCAGCTCACCAACGCAGTTCTCCACCTTCCCGGATGCGTCCCATGTGAGAATGA
SEO ID NO 33 30 32	Consensus  Isolate T8  DK8  SW3  DK11	123 123	CAATGCACCTGCCAGCTCGGATACAAGTAACACCTAATGTGGCTCTGAAACACCGLGGC
30-33 <u>SEO ID NO:</u> 33 30 32 31	CONSENSUS  Isolate T8 DK8 SW3 DK11	184 184	CAATGCACCCTGC-CTGCTGGATACAAGTgACACCTAATGTGGCTGTGAAACACCGCGGC  GCACT-CACTCACAACCTGCGAACGCACGTGATCGTAATGGCAGCTACGGTTGCT
30-33 SEO ID NO: 33 30 32	consensus	245 245	GCACTEACTCAEAACCTGCGA-CaCA-gTeGATCATEGTAATGGCAGCTACGGTCTGCT  CGGCCTTGTATGTGGGGGACGTGTGCGGGGCCGTGATGATAGGGGTCGCAGGCTETCATAAT
31 30-33	DK11 consensus	245	CGGCCTTGTATGTGGGAGACgTGTGCGGGGCCGTGATGATCGTGTCGCAGGCTTTCATAGT

# FIGURE 1D

SEO ID NO	Isolate	
33	T8	306 ATCGCCaGAACGCCACAACTTcACCCAGGGGTGCAACTGTTCCATCTACCAAGGTCATATC
30	DK8	306 ATCGCCLGAACGCCACACTTTACCCAGGAGTGCAACTGTTCCATCTACCAAGGTCATATC
32	SW3	306 ATCGCCAGAACGCCACAACTTTACCCAAGAGTGCAACTGTTCCATCTACCAAGGTCGTATC
31	DK11	306 ATCGCCAGAACACCACTTTACCCAAGAGTGCAACTGTTCCATCTACCAAGGTCacATC
30-33	consensus	ATCGCCaGAACGCCACAACTTLACCCA-GAGTGCAACTGTTCCATCTACCAAGGTCaLATC
SEO ID NO:	Toolato	
33	Isolate T8	367 ACCGGCCACCGCATGGCATGGGACATGATGCTGAACTGGTCACCAACTCTCACCATGATCC
30	DK8	367 ACCGGCCACCGCATGGCATGGGACATGATGCTAAACTGGTCACCAACTCTTACCATGATCC
32	SW3	367 ACCGGCCACCGCATGGCGTGGGACATGATGCTAAACTGGTCACCAACTCTTACCATGATCC
31	DK11	367 ACCGGCCACCGCATGGCATGGGACATGATGCTLAACTGGTCACCAACTCTCACCATGATCC
30-33	consensus	ACCGGCCACCGCATGGCATGGGACATGATGCTAAACTGGTCACCAACTCT-ACCATGATCC
400 TD 110		
SEO ID NO:	<u>Isolate</u> T8	428 TCGCCTAcGCtGCTCGTGTGCCTGAaCTAGtCCTtgAaGTTGTCTTCGGCGGCCATTGGGG
30	DK8	428 TCGCCTATGCCGCTCGTGTTCCTGAGCTAGcCCTccAgGTTGTCTTCGGCGGCCATTGGGG
32	SW3	428 TEGCCTATGCCGCTCGTGTTCCTGAGCTAGTCCTTGAAGTTGTCTTCGGCGGCCATTGGGG
31	DK11	428 TcGCCTATGCCGCcCGTGTTCCTGAGCTAGTCCTTGAAGTcGTCTTCGGtGGtCATTGGGG
30-33	consensus	Tegectaegeegeetegtgtecetgagetageectegaagtegtetteggeggecattgggg
070 TR 110		
SEO ID NO:	<u>Isolate</u> T8	489 CGTGGTGTTTGGCTTGGCCTATTTCTCCATGCAAGGGGGCGGGC
30	DK8	489 CGTGGTGTTTGGCTTGCCTATTTCTCCATGCAGGGAGCGTGGGCCAAAGTCATTGCCATC
32	SW3	489 CGTGGTGTTTGGCTTGGCCTATTTCTCCATGCAAGGAGCGTGGGCCAAGGTCATTGCCATC
31	DK11	489 LGTGGTGTTTGGCTTGGCCTATTTCTCCATGCAGGGAGCGTGGGCCAAGGTCATTGCCATC
30-33	consensus	cgtggtgtttggcttggcctatttctccatgca-ggagcgtgggccaa-gtcattgccatc
686 TR NO		
SEO ID NO:	<u>Isolate</u> T8	550 CTCCTCCTTGTCGCAGGAGTGGACGCA
30	DK8	550 CTCCTtCTTGTCGCAGGAGTGGATGCA
32	SW3	
31	DK11	
30-33	consensus	CTCCT+CTTGTcGCAGGAGTGGA+GCA

### FIGURE 1E

SEO ID	NO: Isolate		
35	DK12	1 tTAGAGTGGCGGAATGTGTCCGGCCTCTACGTCCTTACCAACGACTGTtcCAATAGCAGTA	
36	HK10	1 CTAGAGTGGCGGAATGTGTCTGGCCTCTATGTCCTTACCAACGACTGTCCCAATAGCAGT	ï
37	S2	1 CTAGAGTGGCGGAATACGTCTGGCCTCTATGTCTCAACCAAC	Ä
39	S54	1 CTAGAGTGGCGGAATACGTCTGGCCTCTATaTCCTTACCAACGACTGTTCCAATAGCAGTF	A.
38	S52	1 CTAGAGTGGCGGAATACGTCTGGCCTCTATGTCCTTACCAACGACTGTTCCAATAGCAGT	A.
35-39	consensus	ctagagtggcggaatacgtctggcctctatgtccttaccaacgactgttccaatagcagta	4
SEQ ID	NO: Isolate		
33	DK12	62 Tegtgtatgaggccgatgacgtcattctgcacacacctggctgtgtaccttgtgttcagga	À.
36	HK10	62 TTGTGTATGAGGCCGATGACGTCATTCTGCACACACCTGGCTGTGTACCTTGTGTTCAGGA	1
37	S2	62 TIGIGTATGAGGCCGATGACGTLATTCTGCACACACCTGGCTGTGTACCTTGTGTTCAGG	
39	S54		
		62 TTGTGTATGAGGCCGATGACGTCATTCTGCACACACCCGGCTGTGTACCTTGTGTTCAGGA	L
38	S52	62 TTGTGTATGAGGCCGATGACGTCATTCTGCACACCCCGGCTGTGTACCTTGTGTTCAGGA	
35-39	consensus	TEGTGTATGAGGCCGATGACGTCATTCTGCACACCCEGGCTGTGTACCTTGTGTTCAGGA	
SEO ID			
35	DK12	123 CGGCAATACATCLACGTGCTGGACCTCaGTGACGCCTACAGTGGCAGTCAGGTACGTCGGA	
36	HK10	123 CGGCAATACATCCACGTGCTGGACCCTCGGTGACACCTACAGTGGCAGTCAGGTACGTCGGA	
37	S2	123 CGGtAATACATCCACGTGCTGGACCCCAGTGACACTTACAGTGACGACGACGACGACGACGACGACGACGACGACGACGACG	
39	S54	123 CGGCAATACATCCACGTGCTGGACCCCAGTGACACCTACGGTGACAGTTAAGAGTTAAGGTTAAG	
38	S52	123 CGGCAATACATCCAtGTGCTGGACCCCAGTGACACCTACGGTGGCAGTCAGGTACGTCGGA	
		COOCATTACATECATCOTOCTOCACCCCAGTGACACCTACGGTGGCAGTCAGGTACGTCGGA	
35-39	consensus	CGGCAATACATCCACGTGCTGGACCCCAGTGACACCTACAGTGGCAGTCAGGTACGTCGGA	
SEO ID	NO: Isolate DK12	104 (0) 200 (0) (0)	
33	DKIZ	184 GCAACCACCGCtTCGATACGCAGTCATGTGGACCTGcTAGTGGGCGCGGCCACGATGTGCT	
36	HK10	184 GCAACCACCGC=TCGATACGCAGTCATGTGGACCTGTTAGTGGGCGCGGCCACGATGTGCT	
37	S2	184 GCAACCACCGCTTCGATACGCAGTCATGTGGACCTATTGGTGGGCGCGGCCACLATGTGCT	
39	S54	184 GCAACCACCGCTTCGATACGCAGTCATGTGGACCTATTAGTGGGGGGGG	
38	S52	184 GCAACCACCGCTTCGATACGCAGTCATGTGGACCTATTAGTGGGCGCGGCCACGCTGTGCT	
35-39	consensus	GCAACCACCGCtTCGATACGCAGTCATGTGGACCTatTaGTGGGCGCGGCCACgaTGTGCT	

# FIGURE 1E

SEO ID	NO: Isolate	245	CTGCGCTCTACGTGGGLGATGTGTGGGGCCGTCTTCCTLGTGGGACAAGCCTTCACGTT
36	HK10		CTGCGCTCTACGTGGGCGATATGTGTGGGGCCGTCTTCCTCGTGGGACAAGCCTTCACGTT
37	52		
-			CTGCGCTCTACGTGGGTGATATGTGTGGGGCCGTCTTTCTCGTGGGACAAGCCTTCACGTT
39	S54	245	CTGCGCTCTATGTGGGTGATATGTGTGGGGCCGTCTTTCTCGTGGGACAAGCCTTCACGTT
38	S52	245	CTGCGCTCTATGTGGGTGATATGTGTGGGGCCGTCTTTCTCGTGGGACAAGCCTTCACGTT
35-39	consensus		$\tt CTGCGCTCTAcGTGGGGLGATaTGTGTGGGGCCGTCTTtCTcGTGGGACAAGCCTTCACGTT$
SEO ID	NO: Isolate DK12	306	CAGACCECGTCGCCATCAAACAGTCCAGACCTGTAACTGCTCGCTGTACCCAGGCCAECTT
36	HK10		CAGACCGCGTCGCCATCAAACGGTCCAGACCTGTAACTGCTCGCTGTACCCAGGCCACCTT
37	S2	306	CAGACCTCGTCGCCATCAAACGGTCCAGACCTGTAACTGCTCGCTGTACCCAGGCCATCTT
39	S54	306	CAGACCTCGTCGCCATCAAACGGTCCAGACCTGTAACTGCTCGCTGTACCCAGGCCATCTT
38	S52		CAGACCTCGTCGCCATCAAACGGTCCAGACCTGTAACTGCTCGCTGTACCCAGGCCATgTT
35-39	consensus		CAGACCECGTCGCCATCAAACGGTCCAGACCTGTAACTGCTCGCTGTACCCAGGCCAECTT
SEO ID	NO: Isolate DK12	367	TCAGGACATCGAATGGCTTGGGATATGATGATGATTGGTCCCCCGCLGTGGGTATGGTGG
36	HK10		
			TCAGGACATCGAATGCTTGGGATATGATGATTGCTCCCCCGCcGTGGGTATGGTGG
37	S2	367	TCAGGACATCGCATGGCTTGGGATATGATGATTGGTCCCCCGCTGTGGGTATGGTGG
39	S54	367	TCAGGACATCGAATGGCTTGGGATATGATGATGATGATTGGTCCCCCGCTGTGGGTATGGTGG
38	S52	367	TCAGGACATCGAATGGCTTGGGATATGATGATGAATTGGTCCCCCGCTGTGGGTATGGTGG
35-39	consensus		TCAGGACATCGAATGGCTTGGGATATGATGATGATTGGTCCCCCGCLGTGGGTATGGTGG
SEO ID	NO: Isolate DK12	428	TAGCGCACGTCCTGCGLcTGCCCCAGACCTTGTTCGACATAATAGCLGGGGCCCATTGGGG
36	HK10		1
37	S2		TGGCGCACGTCCTGCGGTTGCCCCAGACCTTGTTCGACATAATAGCCGGGGCCCATTGGGG
			TGCGCACGTECTGCGETTGCCCCAGACCGTGTTCGACATAATAGCCGGGGCCCATTGGGG
39	S54	428	TGCGCACATCCTGCGATTGCCCCAGACCTTGTTTGACATACTGGCCGGGGCCCATTGGGG
38	S52	428	TGGCGCACATCCTGCGATTGCCCCAGACCTTGTTTGACATACTGGCCGGGGCCCATTGGGG
35-39	consensus		TgGCGCACgTcCTGCG-tTGCCCCAGACCtTGTTcGACATAaTaGCcGGGGCCCATTGGGG

## FIGURE 1E

SEO ID NO	: Isolate		
35	DK12	489	CATCATGGCGGGCCTAGCCTATTACTCCATGCAGGGCCAACTGGGCCAAGGTCGCTATCATC
36	HK10	489	CATCTTGGCaGGCCTAGCCTATTACTCCATGCAGGGCCAACTGGGCCAAGGTCGCTATCATC
37	S2	489	CATCTTGGCGGGCCTAGCCTATTACTCCATGCAAGGCAACTGGGCCAAGGTCGCTATCATC
39	S54		
33	554	489	CATCTTGGCGGGCCTAGCCTATTATTCTATGCAGGGCAACTGGGCCAAGGTCGCTATCATC
38	S52	400	COTOTOTOCCCCCCCTACCCCTACCCCTACCCCCTACCCCCCCC
		403	CATCTTGGCGGGCCTAGCCTATTATTCTATGCAGGGCCAACTGGGCCAAGGTCGCTATtgTC
35-39	consensus		CATCHTGGCgGGCCTAGCCTATTACTCCATGCAGGGCCAACTGGGCCAAGGTCGCTATCATC
			The second contract carried ages can cross contract care
SEO ID NO			
SEO ID NO	: <u>Isolate</u> DK12	550	ATGGTTATGTTTTCAGGAGTCGATGCC
35	DK12		
			ATGGTTATGTTTTCAGGaGTCGATGCC
35 36	DK12 HK10	550	
35	DK12	550	
35 36 37	DK12 HK10 S2	550 550	
35 36	DK12 HK10	550 550	
35 36 37	DK12 HK10 S2	550 550 550	
35 36 37 39	DK12 HK10 S2 S54	550 550 550	

# FIGURE 1F

SEO ID NO:         Isolate           43         27           42         26           42-43 consensus (26)	1 GTCAACTATCACAATGCCTCGGGCGTCTATCACATCACCAACGACTGCCCGAACTCGAGCA
SEO ID NO:         Isolate           43         Z7           42         Z6           42-43 consensus (Z6)	62 TAATGTATGAGGCCGAACACCACATCCTACACCTCCCAGGGTGCGTACCCTGTGTGAGGGA
SEO ID NO:         Isolate           43         27           42         26           42-43 consensus (Z6)	123 gGGGAACCAGTCACGCTGCTGGGTGGCCCTTACTCCCACCGTGGCGGGGGCCTTATATCGGT
SEO ID NO:         Isolate           43         27           42         26           42-43 consensus (Z6)	184 GCaccgcTTGAaTCCaTCCGGAGACATGTGGACCTGATGGTAGGCGCLGCTACAGTGTGCT
SEO ID NO:         Isolate           43         Z7           42         Z6           42-43 consensus (Z6)	245 CcGCtCTCTACaTTGGGGACCTGTGCGGTGGcGtATTtTTGGTTGGtCAGATGTTTtTCTTT
SEO ID NO:	306 CCAGCCGCGACGCCACTGGACTACGCAGGACTGCAATTGTTCCATCTALGCGGGGCACGTT
SEO ID NO:   Isolate   43   Z7     42   Z6     42-43   Consensus (Z6)	367 ACAGGCCACAGAATGGCATGGGACATGATGATGAGTCCCACAACCACCLTGGTCC

## FIGURE 1F

43 27 42 26 42-43 consensus (Z6)	428 TCGCCCAGGTLATGAGGATCCCTAGCACTCTGGTgGACCTACTCACTGGAGGGCACTGGGC
SEO ID NO: Isolate   43   Z7     42   Z6     42-43   Consensus (Z6)	489 taTCCTTaTcGGGGTGGCATACTTCtGCATGCAAGCTAATTGGGCCAAGGTCATtcTGGTC
SEO ID NO: Isolate 43 27 42 26 42-43 consensus (Z6)	550 CTTTTCCTCTACGCTGGAGTTGATGCC

## FIGURE 1G

SEO ID NO	: Isolate		
45	SA1	1	GTECCCTACCGGAATGCCTCTGGGGTTTAcCATGTCACCAATGACTGCCCAAACTCCTCCA
47	SA5		GTCCCCTACCGAAATGCCTCTGGGGTTTATCATGTCACCAATGATTGCCCAAACTCTTCCA
49	SA7	1	GTCCCCTACCGAAATGCCTCCGGGGTTTATCATGTCACCAATGATTGCCCGAACTCTTCCA
46	SA4	1	GTTCCCTACCGAAAcGCCTCTGGGGTTTATCATGTCACCAATGATTGCCCAAACTCTTCCA
50	SA13	1	GTTCCCTACCGAAATGCCTCTGGGGTTTATCATGTCACCAATGATTGCCCAAACTCTTCCA
48	SA6	1	GTTCCETACCGGAATGCCTCTGGGGTgTATCATGTEACCAATGATTGCCCCAAACTCTTCCA
45-50	consensus		${\tt GTECCcTACCGAAAEGCCTCEGGGGTETAECATGACCAATGAETGCCCAAACTCETCCA}$
SEO ID NO:	Isolate		
45	SA1	62	TAGTCTACGAGGCTGATAGCCTGATctTGCACGCACCTGGCTGCCTGTGTCAGGCA
			_
47	SA5	62	TAGTCTACGAGGCTGATAACCTGATtCTGCACGCACCTGGTTGCGTGCCCTGTGTCAAGGA
49	SA7	62	TAGTCTALGAGGCTGACAACCTGATCCTGCACGCACCTGGTTGCGTGCCCTGTGTCAGaCA
46	SA4	62	TAGTETACGAGGCTGATAACCTGATCTTGCAEGCACCTGGTTGCGTGCCETGTGTCAGGCA
50	SA13	62	TeGTCTACGAGGCTGATGACCTGATCTTACACGCACCTGGTTGCGTGCCCTGTGTLAGGCA
4.8	SA6		
*0	SA6	62	TAGTCTALGAGGCTGATGACCTGATCCTACACGCACCTGGCTGCGTGCCCTGTGTCCGGAA
45-50	consensus		${\tt TagTcTAcGAGGCTGAtaaCCTGATc-TgCAcGCACCTGGtTGCGTGCCcTGTGTcaggcA}$
SEO ID NO:	<u>Isolate</u> SA1		30
43	SAI	123	AGATAATGTCAGTAGGTGCTGGGTCCAAATCACCCCCCACACTGTCAGCCCCCGACCLTCGGA
47	SA5	123	AGGTAATGTCAGTAGGTGCTGGGTCCAAATCACCCCCACATTGTCAGCCCCGAACCTCGGA
49	SA7	123	AAATAATGTCAGTAGGTGCTGGGTCCAAATCACCCCCACATTGTCAGCCCCGAACCTCGGA
46	SA4	123	AGATAATGTCAGTAAGTGCTGGGTCCAAATCACCCCCACGTTGTCAGCCCCGAALCTCGGA
50	SA13		GGGTAATGTCAGTAGGTGCTGGGTCCAGATCACCCCCACACTGTCAGCCCCGAGCCTCGGA
48	SA6		GGATAATGTCAGTAGATGCTGGGTtCAtATCACCCCCACACTATCAGCCCCGAGCCTCGGA
45-50	consensus		agaTAATGTCAGTAggTGCTGGGTcCAaATCACCCCCACa-ToTCAGCCCCGAaccTCGCA

FIGURE 1G

SEO ID NO	: <u>Isolate</u> SA1	184 GCGGTCACGGCTCCTCTTCGGAGGGCCGTTGACTACTTAGCGGGAGGAGCTGCtCTCTGC1
47	SA5	184 GCGGTCACGGCTCCTCTTCGGAGGGtCGTTGACTACTTAGCGGGAGGGGCTGCCCTCTGCT
49	SA7	184 GCGGTCACGGCTCCTCTTCGGAGGGCCGTTGACTACCTAGCGGGAGGGCTGCCCTCTGCT
46	SA4	184 GCGGTCACGGCTCCTCTTCGGAGGGCCGTTGACTTAGCGGGAGGGGCTGCCCTCTGCT
50	SA13	184 GCGGTCACGGCTCCTCTCGGAGGGCCGTTGACTTAGCGGGGGGGG
48	SA6	184 GCGGTCACGGCTCCTCTCGGAGGGCCGTTGALTACTTgGCGGGAGGGCCGCCCTgTGCT
45-50	consensus	GCGGTCACGGCTCCTCTTCGGAGGGcCGTTGAcTACtTaGCGGGaGGGCtGCcCTcTGCT
SEO ID NO	<u> Isolate</u>	
45	SA1	245 CCGCACTATACGTCGGcGACGCGTGCGGGGCAGTGTTtccTGGTAGGCCAAATGTTCACCTA
47	SA5	245 CCGCACTATACGTCGGGGACGCGTGCGGGGCAGTGTTCTTGGTAGGCCAAATGTTCACCTA
49	SA7	245 CCGCgCTATACGTCGGGGACGCGTGCGGGGCAGTGTTTTTGGTAGGCCAgATGTTCAGCTA
46	SA4	245 CCGCaCTATACGTCGGGGACGCGTGCGGGCAGTGTTTTTGGTAGGCCAAATGTTCACCTA
50	SA13	245 CCGCGTTATACGTCGGAGACGCCGTGCGGGGCAGTGTTTTTGGTAGGtCAAATGTTCACCTA
48	SA6	245 CCGCGTTATACGTCGGAGACGLGTGCGGGGCALTGTTTTTGGTAGGcCAAATGTTCACCTA
45-50	consensus	CCGC-CTATACGTCGGGGACGCGTGCGGGGCAGTGTTttTGGTAGGcCAAATGTTCAcCTA
SEO ID NO:	Isolate	
45	SA1	306 TAGGCCTCGCCAGCATACCACAGTGCAGGACTGCAACTGTTCCATTTACAGtGGCCATATC
47	SA5	306 TAGGGCTCGCCAGCATACTACGGTGCAGGACTGCAACTGTTCCATTTACAGCGGCCATATC
49	SA7	306 TAGGCCTCGCCAGCACTACGGTGCAGGACTGCTACTGTTCCATTTACAGTGGCCATATC
46	SA4	306 TAGGCCTCGCCAGCACTACGGTGCAAGACTGCAACTGCTCLATTTACAGTGGCCATATC
50	SA13	306 TAGCCCTCGCCgGCATAATgttGTGCAGGACTGCCAACTGtTCCATTTACAGTGGCCACATC
48	SA6	
45-50	consensus	TAGGCCTCGCCAGCAtactacgGTgCAgGACTGCAAcTGtTCcATTTACAGtGGCCAtATC
SEO ID NO:	Isolate	
45	SA1	367 ACCGGCCACCGGATGGCLTGGGACATGATGATGATTGGTCACCTACGACAGCCTTGcTGA
47	SA5	367 ACCGGCCACCGAATGGCATGGGACATGATGATTAGTCACCTACGACAGCCTTGGTGA
49	SA7	367 ACCGGCCACCGAATGGCATGGGACATGATGATTGGTCACCTTACGACACCCTTACGTCA
46	SA4	367 ACCGGCCACCGGATGGCATGGCATGATGATTGGTCACCTACCT
50	SA13	367 ACCGGCCACCGGATGGCATGGGACATGATGATGATTGGTCACCTACAACAGCTTTGGTGA
48	SA6	367 ACLEGGCCACCGGATGGCATGGACATGATGATTGGTCACCCGCGACAGCCTTGGTGA
45-50	consensus	ACCGGCCACCGGATGGCATGGGACATGATGATTGGTCACCtaCgACaGCcTTGgTGA

84892\_

FIGURE 1G

\$\frac{\text{SEO ID NO}}{45}\$ 47 49 46 50 48	SA1 SA5 SA7 SA4 SA13	428 TGGCCCAGATGCTACGGATCCCCAGGTGGTCATAGACATCATAGCCGGGGGCCACTGGGG
45-50	consensus	TGGCCCAgtTGcTACGGATtCCCCAggTGCTCATtCAGATGATT
SEO ID NO:		TGGCCCAgtTGcTACGGATtCCCCAgGTGGTCATtGACATCATtGCCGGGGgCCACTGGGG
45	SA1	4.9.9. (CONCURRENCE COCCURRENCE
	JAI	489 GGTCTTGTTtGCCGcCGCATACTTtGCGTCgGCcGCcAACTGGGCTAAGGTaGTGCTGGTt
47	SA5	489 GGTCTTGTTCGCCGLCGCATACTTCGCGTCAGCCGGCTAACTGGGCTAAGGTTGTGCTGGTC
49	SA7	489 GGTCTTGTTCGCCGCCGCATATTTCGCGTCAGCGGCTAACTGGGCTAAGGTTGTGCTGGTC
46	SA4	489 GGTCTTGTTLGCCGCCGCATATTTCGCGTCAGCGGCTAACTGGGCTAAGGTTaTaCTCGTC
50	SA13	489 GGTCTTGTTCGCCGCCGCATACTaCGCGGGGCTAACTGGGCAAAGGTTGTGCTGGTC
48	SA6	489 GGTCTTGTTCGCCGCtGCATACTtCGCGTCGGCGGCTAACTGGGCLAAGGTTGTGCTGGTC
45-50	consensus	GGTCTTGTT-GGCCG-cGCATAcTt-cGCGTC-GCGGCLAACTGGGCLAAGGTLGTGCTC
SEQ ID NO:	<u>Isolate</u>	
45	SA1	550 CTCTTTCCTCTTTTCCCCCCCCCCCCCCCCCCCCCC
	SAI	550 CTGTTcCTGTTTGCGGGGGTCGATGGC
47	SA5	550 CIGTITICTGTTGCGGGGGCTCGATGGC
49	SA7	550 TTGTTTCTGTTTGCGGGGGTCGATGCC
46	SA4	550 TIGTITCTGTTTGCGGGGGTCGATGCC
50	SA13	550 CTGTTTCCGGGGGGTCGATGCC
48	SA6	550 tTGTTTCTGTTTGCGGGGGTtGATGCC

-TGTTLCTGTTTGCGGGGGTcGATGcC

45-50

consensus

SEO ID NO: Genotype (IV/2b)

### FIGURE 1H

SEC ID NO:								
30-33	(IV/2b)	1	GTGGAAGTCAGGAACAtCAC	TTCt AGCTACTA	CCCCACCA	ATC ATT		~-
34	(2c)	1	GTGGAGGTCAAGGACACCG	CCACTCCTACAT	GCCGACCA	VCC VIIII	SCICALACARCAGO	ĽA.
26-29	(III/2a)	1	GCCCAAGTGAAGAACACCAG	taccaceThen	CCTCACCA	ACGAII(	SCICCAACICTAG.	ľA
35-39	(V/3a)	. 1	CTAGAGTGGCGGAATacGTC	***********	GGTGACCA	ACGACTO	JETCCAAEGACAG	CΑ
9-25	(II/1b)		t At GA a GTG Co CA A CGT or TO	CCCCC	CCTEACCA	ACGACTO	TECCAATAGCAG:	ΓA
1-8	(I/la)	1	tAtGAaGTGCGCAACGTgTC	-COGGGTGTACCA	tGTCACGA	ACGACTO	GCTCCAACTcaAG	ca
40	(4a)	ī	tACCAAGTGCGCAACTCcaC	GGGGCTTTACCA	TGTCACCA	ATGALT	SCCCTAAcTCGAG	ĽΑ
42-43	(4c)	i	GAGCACTACCGGAATGCTTC	GGGCATCTATCA	CATCACCA	ATGATT	TCCGAATTCCAG	ΓA
44	(4d)		GTTAACTATCGCAATGCCTC	GGGCGTCTATCA	CGTCACCA	CGACT	CCCGAACTCGAG	ZΑ
41	(4b)	- 1	TACAACTATCGCAACAGCTC	GGGTGTCTACCA	TGTCACCA	CGATT	SCCCGAACTCGAG	CA
45-50	(5a)		GTGCACTACCGGAATGCTTC	GGGCGTCTATCA	TGTCACCA	TGATTO	CCCTAACACCAG	CA
51	(6a)		GTtCCcTACCGaAAtGCCTC	tGGGGTtTAtCA	TGTcaccai	TGALTO	CCCAAACTCLTCC	ZA.
31	(64)	1	CTTACCTACGGCAACTCCAG	TGGGCTATACCA	TCTCACAA	TGATTO	CCCCAACTCCAG	CA
1-51								
1-21	consensus		A	TA	AC AZ	GA TO	CAA	
SEO ID NO:	Genotype							
30-33	(IV/2b)	62	TCACCTGGCAACTCACCAAC	GCAGTt CTCCAC	CTTCCCGG	TGCGTC	CCattettenenant	270
34	(2c)	62	TCGTTTGGCAGCTTGAAGGA	المالا كالمصاحب المالي الإساكا	لا ت محمد المحمد الم	TOOMO		_
26-29	(III/2a)	62	TUACCTUGUAACTCCAGGCC	GCGGTcCTCCAC	TOCOCCC	TC+~TC	CCCTC-CACA	
35-39	(V/3a)							
9-25	(II/1b)	62	TtGTGTatGAggCAgcgGAC	aTGATCaTGCAC	ACCCCCCCC	TOTOTA	CCCTTGTGTTCAGG	xA.
1-8	(I/la)							
40	(4a)	62	TAGTCTATGAAGCTGACCAT	CACATCCTACAC	rreceeses	TOIGIC	CCTTGCGTTCGCG	iA.
42-43	(4c)	62	TAGTGTATGAGGCCGAACAC	Chanteethan.	TOCCOGGG	TGCGTA	CCCTGTGTGATGA	rC.
44	(4d)	62	TAGTCTATGAAACCGATTAC	CACATCTTACAC	TOCCAGGG	TGCTTg	CCCTGTGTGAGGG	it
41	(4b)	62	TAGTGTACGAGACGGAGCAC	CACATCATCACAC	TTCCCGGGA	TGCGTT	CCTTGCGTGAGGG	A
45-50	(5a)	62	TagTcTAcGAGGCTGAtaaC	CTCATCATGCAC.	TOCCAGGG	TGTGTC	CCCTGTGTGCGGA	۲C
51	(6a)	62	TCGTGCTGGAGGCGGATGCT	ATTENTO	CACCIGGE	TGCGTG	CCcTGTGTcaggc	:A
			- colociounoucounidei,	AIGAICTIGCAT	FIGCCIGGA	TGCTTG	CCTTGTGTGAGGG	т
1-51	consensus		T A					
				T T CA	CC GG	TG T	CC TG G	
SEO ID NO:	Genotype							
30-33	(IV/2b)	122	CANTICCCA CC-TCC-TCC-TCC					
34	(2c)	123	CAATGGCACCCTGCGCTGCTC	GATACAAGTGAC	ACCTAATG	TGGCTG	TGAAACACCGcGG	·C
26-29	(III/2a)							
35-39	(V/3a)							
9 - 25	(II/1b)							
1-8	(I/1a)							
40	(1/1a) (4a)							
42-43	(4a) (4c)							
44								
41	(4d)							
	(4b)							
45-50	(5a)							
51	(6a)	123	CGATGATCGGTCCACCTGTTC	GCATGCTGTGAC	CCCCACCC	recera	PACCA A ATTCOMIC	~
						. CCCCA	INCUMMICCITIC	_

TG TGG

T C CC A T C

1-51

consensus

### FIGURE 1H

SEO ID NO:	Genotype		
30-33	(IV/2b)	184	GCaCTCACTCACAACCTGCGAaCaCAtgTcGAcaTGATcGTAATGGCAGCTACGGTCTGCT
34	(2c)	184	GCTCTCACTAAGGGCCTGCGAGCACACATCGATATCATCGTGATGTCTGCTACGGTCTGTT
26-29	(III/2a)	184	GCCTCACGCAGGGCTTGCGGACGCACATCGACATGGTLGTGATGTCCGCCACGCTCTGCT
35-39	(V/3a)	184	GCAACCACCGCtTCGATACGCAGTCATGTGGACCTatTaGTGGGCGCGGCCACGATGTGCT
9-25	(II/1b)	184	gTCcCcACtAcGaCaATACGACgcCAcGTCGAtTTGCTCGTTGGGGCGGCTgctTTCTGcT
1-8	(I/la)	184	CTCCCcgCAaCGCAgCTtCGACGTcACATCGAtCTGCTtGTCGGgAGcGCCACCCTCTGcT
40	(4a)	184	GCTCCGCTTGAGTCGTTCCGGCGACATGTGGACTTAATGGTAGGCGCGGCCACTTTGTGTT
42-43	(4c)	. 184	GCtCCGCTTGACTCCCTCCGGAGACATGTGGACCTGATGGTgGGCGCCGCTACtGTaTGCT
44	(4d)	184	GCTCCGCTTGAGTCTTTGAGACGTCACGTGGATCTGATGGTGGGCGCGCGC
41	(4b)	184	GCACCGTTAGAGTCCATGCGCAGGCATGTAGACCTGATGGTGGGTG
45-50	(5 <b>a</b> )	184	GCGGTCACGGCTCCTCTTCGGAGGGCCGTTGACTACtTaGCCGGGGGGGCCCCTCTGCT
51	(6a)	184	ACGCCCGCAACGGGATTCCGCAGGCATGTGGATCTTCTTGCGGGCGCCGCAGTGGTTTGCT
1-51	consensus		T G T GA T G GC T TG T
SEO ID NO:	<u>Genotype</u>		
30-33	(IV/2b)	245	
34	(2c)	245	CIGCCCITIATGIGGGGACGTGTGTGGGCGCGCTGATGCTGGCCGCTCACGTCGTCGTCCT
26-29	(III/2a)	245	CCGCtCTtTACGTGGGGGAccTCTGCGGGGGGGGTGATGCTCGCaGCcCAgATGTTCATtaT
35-39	(V/3a)	245	CIGCGCICTACGTGGGGGGGTGTGTGTGGGGACAAGCCTTCACGTT
9-25	(II/1b)	245	CCGctATGTAcGTGGGGAtCTcTGCGGaTCtGTttTCCTccTcTCCCAGcTGTTCACctT
1-8	(I/la)	245	CGGCCCTCTACGTGGGGGACtTGTGCGGGTCTGTCTTtCTtGTCGGtCAaCTGTTCACCtT
40	(4a)	245	CIGCCTCTATGTTGGGGACCTCTGCGGAGGTGCCTTCCTGATGGGGCAGATGATCACTTT
42-43	(4c)	245	CtGCCCTCTACGTTGGaGAtCTGTGCGGTGGtGCATTCTTGGTTGGCCAGATCTTCTCCTT
44	(4d)	245	CCGCCCTCTACATCGGAGACGTGTGTGGGGGGGTGTTCTTGGTCGGTC
41	(4b)	245	CCGCCTTCTACATTGGAGATCTGTGTGGGGGGGGGCTCTTCCTAGTGGGCCAGCTGTTCGACTTT
45-50	(5a)	245	CCGCGCTATACGTCGGGGACGCGTGCGGGGCAGTGTT++TGGTAGGCCAAAATGTTCAACAA
51	(6a)	245	CATCCCTGTACATCGGGGACCTGTGTGGGCTCTCTCTTTTTGGCGGGGACAACTATTCACCTT
1-51	consensus		C T TA T GG GA TG GG T T CA T
SEO ID NO:			
30-33	Genotype		
30-33	(IV/2b)	306	ATCGCCaGAACGCCACAACTTLACCCAAGGTGCAACTGTTCCATCTACCAAGGTCaLATC
26-29	(2c) (III/2a)	306	GTCGCCACAACACCATACGTTTGTCCAGGAATGCAACTGTTCCATATACCCGGGCCGCATT
35-39	(V/3a)	306	CTCGCCGCaaCacCACTgGTTTGTGCAaGAaTGCAAtTGCTCcATcTACCCtGGtACCATC
9-25	(II/1b)	306	CAGACCECGTCGCCATCAAACGGTCCAGACCTGTAACTGCTCGCTGTACCCAGGCCAECTT
1-8	(I/1a)	306	cTCgCCtCGcCggcAtgaGACagtaCAGgAcTGcAAcTGcTCaaTCTATCCcGGcCacgTa
40	(4a)	306	cTCtCCCAGgCgCcaCTGGACaACGCAaGaCTGcAAtTGTTCtATCTAtCCcGGCCAtATa
42-43	(4c)	306	TCGGCCGCGTCGCCACTGGACCACGCAGGAGTGCAATTGTTCCATCTACACTGGCCATATC
44	(4d)	306	CCAGCCGCGACGCCACTGGACTACGCAGGACTGCAATTGTTCLATCTACGCAGGGCALATC
41	(4b)	306	CCAACCTCGCCGCCACTGGACCACCCAAGACTGCAATTGTTCCATCTACACAGGACATATC
45-50	(5a)	306	CCGACCGCGCGCACTGGACCACCCAGGATTGCAACTGCTCCATCTATCCTGGTCACGTC
51	(6a)	306	TAGGCCTCGCCAGCAtactacgGTgCAGGACTGCAACTGtTCCATTTACAGtGGCCAtATC
1-51	consensus	500	TCAGCCCCGCCGTCATTGGACTGTGCAAGACTGCTACCATCTATACAGGCCACGTC CC C CA TG AA TG TC T TA GG T
			CC C CA TG AA TG TC T TA GG T
SEO ID NO:	Genotype		
30-33	(IV/2b)	367	ACCGGCCACCGCATGGCaTGGGACATGATGCT&AACTGGTCACCAACTCTLACCATGATCC
34	(2c)	367	ACGGGACACCGCATGGCTTGGGATATGATGATGATGTCGCCCCACTACCACCATGCTCC
26-29	(III/2a)	367	ACEGGACACCGTATGGCATGGGACATGATGATGAACTGGTCGCCCACTACCACCATGCTCC ACEGGACACCGTATGGCATGGATGATGATGATCGCCCACGGCCACCACCACCATGCTCC
35-39	(V/3a)	367	TCAGGACATCGAATGGCTTGGGATATGATGATGATTGGTCCCCCGCLGTGGGTATGGTGG
9-25	(II/1b)	36/	TUAGGIUACCGCATGGCTTGGGATATGATGATGATGATGATGATGATGATGATGATG
1-8	(I/la)	367	ACGGGtCACCGcATGGCaTGGGATATGATGATGAACTGGTCCCCtACgaCgGCgCTGGTag
40	(4a)	36/	ACCIGCCACAGGATGGCGTGGGACATGATGATGATGACCCCTACCACCACCACCTACCACCACCACCACCACCACC
42-43	(4c)	36/	ACGGGCCACAGGATGGCATGGATGATGATGACTGGAGTCCCACAACCACCGTCGTC
44	(4d)	36/	ACAGGACACAGAATGGCTTGGGACATGATGATGATTGGAGCCCCA CTCCCA CCCTCCTCC
41	(4b)	36/	TCGGGCCACAGGATGGCCTGGGACATGATGATGAACTGGAGCCCTACCAGCCCCCCCACATTA
45-50	(5a)	36/	ACCOGCCACCGGATGGCATGGCATGATGATGATTGGTCACC+aCaCcaCcaTTCATCA
51	(6a)	367	ACCOGCCACAGGATGGCTTGGGACATGATGAACTGGTCACCCACAACCACTCTCGTCC
1-51	consensus		C GG CA G ATGGC TGGGA ATGATG T AA TGG CC C T T

### FIGURE 1H

SEO ID NO	Genotype																		
30-33	(IV/2b)	428	т	ace	'A+1	aceaa	*+ 007	YOU .		7 n OFF									
34	(2c)	428	TO	CCCT	70	Table Co.	CCC.	יאמרי	2000	AGC I	AGE CO	TtgAa	IGTE	GTC	TTC	GCGC	CCA:	TGGG	G
26-29	(III/2a)	428	TYC	CCCT	יאכי	CCCN	racac		raca	MAGI	CATCI	TGGAT	CATT	GIT.	ACAC	GAGG	TCA:	TGGG	G
35 - 39	(V/3a)	429	To	GCGC	'AC	OTO OT	2000		20000	AGGT	CATCA	TAGAC	CATC	aTt	aGCG	GgGG	CtCA	TGGG	G
9-25	(II/1b)	420	T-9	TCCC		gree!	CCC-			AGAC	CtTGT	TCGAC	CATA	aTa	GCc0	GGGC	CCA	TGGG	G
1-8	(I/la)	420	To	CC+C	AG		C-C	a1(	ccac	AAGC	TgTCg	TGGA	caTG	GTg	gCgG	:GgGC	CCAC	TGGG	G
40	(4a)																		
42-43	(4c)	120	TC		700	TT-N	CAGG	GT	CCCC	CAGC	CTTTC	TCGAC	ATG	GTT	GCCG	GAGG	CCAC	TGGG	G
44	(4d)	420	TC		700	31 CA1	CAGG	ATC	CCT	GCAC	TCTGG	TagAt	CTA	CTC	CTG	GAGG	GCAC	TGGG	G
41	(4b)	420	T-C		70	-11941	IN CCC	AIC	CCAC	GCGC	CATGG	TCGAC	CTG	CIT	GCAG	GCGG	CCAC	TGGG	G
45-50	(5a)	420	70	2000	7	41011	ACGG	ATC	CCC	CTAT	CCTAG	GTGAC	TTG	CTC	ACCG	GGGG	TCAC	TGGG	G
51	(6a)	420	TA	JCCC TCTN	AGI	1001	ACGG	ATT	cccc	AgGT	GGTCA	TtGAC	ATC	ATt(	GCCG	GGGg	CCAC	TGGG	G
	(64)	428	1A	ICIA	نای	ATCTI	GAGG	GTA	CCTC	AGAT	TTGTG	CGAGI	GTG	ATA:	TTG	GTGG	CCAT	TGGG	G
1-51	consensus		т																
	Combeniada		T	С			G	т	CC				T	T	G	GG	CA	TGGG	G
SEO ID NO:	Genotype																		
30-33	(IV/2b)	400	-01		~														
34	(2c)	400	CG.	I GGT	GTI	TGGC	TIGG	CCI	ATTT	CTCC	ATGCA	gGGAG	CGT	GGGC	CAA	aGTC	ATtG	CCAT	c
26-29	(III/2a)	*07	16.	THAT	61.1	1000	CICG		AC-I-L	CTCCI	מתהרמו	ユベこンエ	COTY	2000	777	~~~~	3 maa	~~~	_
35-39	(V/3a)	489	CG.	Can	GTT	CGGC	tTaG	CCT	ACTI	CTCT	ATGCA	GGGAG	CGT	GGGC	GAA	aGTC	qTTG	TCATO	2
9-25	(II/1b)	407	<u></u>		ىس	JUUP	CIAG		ATTA	crcci	ΔΤΥΓΓΔ.	тсссъ	الملت لا	2020	ת תחי	COTO	~~~~	mm	~
1-8	(II/ID) (I/Ia)																		
40	(4a)																		
42-43	(4c)																		
44	(4C) (4d)	407	cg.	CCT	1 d I	CGGG	tTGG(	COT	ACTT	CaGtz	ו עם ייצרע	ע בייים	ATTY	2000	ממסי		Nm-0	maama	~
41	(4b)	407	CA.		201	1000	MIAG	-61	ACTT	CAGCZ	ו ביי אדנ	مصىء	DALA A	2020	ית תחי	~~~~	Nmac	maama	~
45-50	(5a)	489	AGI	TCT.	LGC	TGGT	CTAG	. II.	سلمك	$C\Delta CCD$	מתיכראו	TACTA.	7		~~~	~~~~			_
51	(5a)	407	GGI	CIT	211	CGCC	GCCG	CAT	ACTE	CGCGT	ורתפרי	TCC+ N	M C-TV	2000	- 77	~m~			
31	(6a)	489	GAT	ACT	ACT	AGCC	GTTG	CT.	ACTT	TGGCA	TGGC	rggca.	ACTO	GCT	AAA	AGTT	CTGG	CTGTT	ė
1-51																			
1-21	consensus		1	`Т		G	G	Т	т				TO	G:	AA	GT		т	
SEO ID NO:	C																	-	
30-33	Genotype (IV/2b)																		
34	(1V/2B) (2c)	550	CTC	CTt	TT	GTcG	CAGG	CT	GGAt	<b>GCA</b>									
26-29	(III/2a)	550	CIC	CTG	TG.	ACTG	TTGGC	GTY	GGAG	GCG									
35-39		550	CTt	tTG	тg	gCcG	TTGG	ſGΤ	GGAC	GCG									
9-25	(V/3a)	550	ATG	gTT	TG	TTTT	CAGG	(GT	CGAt	3CC									
1-8	(II/1b)	550	aTG	CTAC	TC	TTTG	CCGG	GT	t GAc	3Gg									
40	(I/la)	550	CTG	tTGC	Tg	TTtg	CCGGC	GT	CGAt	CG									
42-43	(4a)	550	CII	TTCC	TC	TTTG	TGGC	GT	AGAC	GCC									
	(4c)	550	CTT	TTCC	TC	TtCG	TGG	GT.	<b>IGAT</b>	SCC									
44 41	(4d)	550	CTG	TTTC	TC	TTTG	TGGA	GT	CGAC	CT									
	(4b)	550	CTA	TTCC	TC	TTTG	CGGG	GT	CGAGO	GA									
45-50 51	(5a)	550	tTG	TTtC	TG	TTTG	CGGGG	GT	GATO	GCC.									
21	(6a)	550	CTG	TTCC	TA	TTTGC	AGGG	GT	rgaac	CA									

C GG GT GA G

TTT

1-51

consensus

### FIGURE 2A

SEO ID NO:	<u>Isolate</u> S14	
		1 YQVRNSTGLYHVTNDCPNSSIVYELADAILHaPGCVPCVREGNtSRCWVAMTPTVATRDGK
52	DK7	I VARISTED IN VINDEPNSSIVIRAADAILHTPGCVPCVREGNVSRCWVAMTPTVATRDGK
59	US11	1 YOVRNSTGLYHVTNDCPNSSIVYEAADAILHTPGCVPCVREGNASRCWVAMTPTVATRDGK
55	DR4	
54	DR1	
53	DK9	1 YOVENSSGLYHVINDCPNSSIVYFAADAILHSDCGIDGUDGCVACAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
58	SW1	
57	S18	
52-59	consensus	yQVRNStGLYHVTNDCPNSSIVYEaADaILH-PGCVPCVREgnasrCWVavtPTVATRDGK
SEO ID NO:	<u>Isolate</u>	
56	S14	62 LPatQLRRyIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRR1WTTQdCNCSIYPGHI
52	DK7	62 LPTaQLRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRRHWTTQGCNCSIYPGHI
59	US11	62 LPTTQLRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRRHWTTQGCNCSIYPGHI
55	DR4	62 LPTTQLRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRhHWTTQDCNCSIYPGHI
54	DR1	62 LPTTOLRRHIDLLVGSATLCSALYVGDLCGSVELVGOLFTESDEDUNTTODCVGSTVERU
53	DK9	62 LPATOLRRHIDLIVGSATICSALVIGNI COSVETUCO ETRICADA
58	SW1	62 LPATOLRRHIDLLVGSATLCSALYVGDLCGSVELVSOLETESPREUMTTODCSVGGTVDGVT
57	S18	62 LPATQLRRHIDLLVGSATLCSALYVGDLCGSVFLVSQLFTiSPRRHWTTQDCNCSIYPGHI
52-59	consensus	LP-tQLRRhIDLLVGSATLCSALYVGDLCGSVFLVgQLFTfSPRrhWTTQdCNCSIYPGHI
SEO ID NO:	Isolate	
56	S14	123 TGHRMAWDMMNWSPTTALVVAQLLRIPQAILDMIAGAHWGVLAGIAYFSMVGNWAKVLVV
52	DK7	123 TGHRMAWDMMNWSPTTALWVAOLLBIDOATIDMIACAWWGU ACIAWRGU
59	US11	123 TGHRMAWDMMNWSPTaALVVAOLLRIPOATLDWIAGAHWGULAGIAVEGMYCNYAGIA
55	DR4	123 TGHRMAWDMMNWSPTTALVVAOLLRIPOATLDMTAGAHWGVI AGTAYRSMYCNTANG UN
54	DR1	123 TGHRMAWDMMNWSDTTALIAMAOLI BIROATI DATACALLA CALLACTICAL CALLAC
53	DK9	123 TGHRMAWDMMNWSPTaALVMAOLLRIPOATLDMTAGAHWGUI AGTAYESMIGNIA HANNA
58	SW1	123 TGHRMAWDMMMNWSPTTAT.VvAOLI.PTDOAYI.DMTAGAYWGU AGAYYGU
57	S18	
52-59	consensus	TGHRMAWDMMMNWSPTLALVVAQLLRiPQAiLDMIAGAHWGVLAGIAYFSMvGNWAKVlvV

SEO ID NO	<u> Isolate</u>	
56	S14	184 LLLFAGVDA
52	DK7	184 LLLFAGVDA
59	US11	184 LLLFAGVDA
55	DR4	184 LLLFAGVDA
54	DR1	184 LLLFAGVDA
53	DK9	184 LLLFtGVDA
58	SW1	184 LLLFsGVDA
57	S18	184 LLLFaGVDA
52-59	consensus	LLLFaGVDA

# FIGURE 2B

1	SEO ID NO:	Isolate		
	75	T10	1	YEVRNVSGmYHVTNDCSNSSIVfEAaDlIMHTPGCVPCVREgNsSRCWVALTPTLAARNtS
	62	DK1	1	YEVRNVSGVYHVTNDCSNSSIVYEAVDVIMHTPGCVPCVRENNhSRCWVALTPTLAARNAS
	64	UVA	1	held-wiss i vid time service times a part of the service times and
	09	nn-	1	II III IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
	76	US6	1	YEVENUSGMYHUTINDCSNSSTUYED ADMINISTRACTION OF COUNTRY TO MILE OF COUNT
	-		_	
	68	IND8	1	YEVRNVSGVYHVTNDCSNSSIVYEAADMIMHTPGCVPCVREGNfSsCWVALTPTLAARNAS
	67	IND5	1	YEVRNVSGVYHVTNDCSNSSIVYEAADMIMHTPGCVPCVREGNSSRCWVALTPTLAARNAS
	72	CMO		
	73	342	1	1EVRNVSGV1HV1NDCSNSS1V1ETADM1MHTPGCVPCVREANSSRCWVALTPTLAARNES
	63	HK3	1	YEVENIVSGTYHVINDCSNSSWIVSETADMIMUTDGGVDGADENNSSEDGWAA A DANG
	••		-	
	66	HK8	1	YEVRNVSGIYHVTNDCSNSSIVYETADMIMHTPGCmPCVRENNSSRCWVALTPTLAARNVS
	61	D3	1	YEVRNVSGVYQVTNDCSNSSIVYETADMIMHTPGCVPCVREdNSSRCWVALTPTLAARNSS
71 S45 1 YEVRNYSGYHVINDCSNSIVYEAVDWIHTPGCVPCVERINSSRCWALTPTLAARNS 72 SA10 1 YEVRNYSGHVINDCSNSIVYEAVDWIHTPGCVPCVERINSSRCWALTPTLAARNS 69 P10 1 YEVRNYSGHVINDCSNSSIVYEADDMIMHTPGCVPCVERINSSRCWALTPTLAARNS 60 D1 1 YEVRNYSGVYHVINDCSNSSIVYEADDMIMHTPGCVPCVERINSSRCWALTPTLAARNS 61	74	Т3	1	YEVRNVSGVYyVTNDCSNSSIVYETADMIMHTPGCVPCVREsNSSRCWVALTPTLAARNAS
71 S45 1 YEVRNYSGYHVINDCSNSIVYEAVDWIHTPGCVPCVERINSSRCWALTPTLAARNS 72 SA10 1 YEVRNYSGHVINDCSNSIVYEAVDWIHTPGCVPCVERINSSRCWALTPTLAARNS 69 P10 1 YEVRNYSGHVINDCSNSSIVYEADDMIMHTPGCVPCVERINSSRCWALTPTLAARNS 60 D1 1 YEVRNYSGVYHVINDCSNSSIVYEADDMIMHTPGCVPCVERINSSRCWALTPTLAARNS 61	65	HK2	1	VEVENIVE CONTRACTOR OF THE TABLET CONTRACTOR OF CHARLES AND ASSESSMENT OF THE ASSESS
72 SA10 1 YEVRNVSGWYHVINDCSNSSIVYEAADMIHHTPGCVPCVRENNSSRCWVALTPILARRNS 69 P10 1 YEVRNVSGWYHVINDCSNSSIVYEAADMIHHTPGCVPCVRENNSSRCWVALTPILARRNS 60 D1 1 YEVRNVSGWYHVINDCSNSSIVYEADMIHHTPGCVPCVRENSSRCWVALTPILARRNS 60 D1 1 YEVRNVSGWYHVINDCSNSSIVYEADMIHHTPGCVPCVREDSSRCWVALTPILARRNGN 60 D1 1 YEVRNVSGWYHVINDCSNSSIVYEADMIHHTPGCVPCVREDSSRCWVALTPILARRNGN	03		-	
72 SA10 1 YEVRNYSGYHVINDCSNSIVYEADMIMHTPGCVPCVEGMSSGCWALTPTLAARNS 69 P10 1 YEVRNYSGYHVINDCSNSIVYEADMIMHTPGCVPCVVEGMSSGCWALTPTLAARNS 60 D1 1 YEVRNYSGYHVINDCSNSIVYEADMIMHTPGCVPCVVEGMSSGCWALTPTLAARNS 60 IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	71	S45	1	YEVRNVSGaYHVTNDCSNSSIVYEAvDvI1HTPGCVPCVPENNSSPCWVALTDTLAADNSS
69 P10 1 YEVRNVSCVYHVTNDCSNSSIVYEADMIMHTPGCVPCVERMSSCWVALTPILARMSS				
60 D1 1 YEVRAVSCVINVINDCSINSILVIELADMIMHTPGCVPCVEEASSSCWMALTPTLAARNIGN	72	SA10	1	YEVRNVSGmYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNSS
60 D1 1 YEVRAVSCVINVINDCSINSILVIELADMIMHTPGCVPCVEEASSCWMALTPTLAARNIGN			_	
	69	P10	1	YEVRNVSGVYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNSS
	60	D1	1	VEVENUS CVVHVTNDC SNS STAVE + DMINGTED COLDCING ANG GROWING ARROWS
70 S9 1 YEVRNVSGAYHVTNDCSNSSIVYEAADvIMHTPGCVPCVqEgNSSqCWVALTPTLAARNAt	•••		-	
	70	S9	1	YEVRNVSGaYHVTNDCSNSSIVYEaADvIMHTPGCVPCVgEgNSSgCWVAI.TPTI.AAPNat
60-76 consensus yEVrNVSGvYhVTNDCSNsSiVyEaaDmImHTPGCvPCVrEnNsSrCWVALtPTLAARNas	60-76	consensus		yEVrNVSGvYhVTNDCSNsSiVyEaaDmImHTPGCvPCVrEnNsSrCWVALtPTLAARNas

FIGURE 2B

SEO ID NO:	Isolate		
75	T10	62	vPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSOLFTFSPRRHET1ODCNCSIYPGH1
-			
62	DK1	62	IPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETaQDCNCSIYPGHV
64	HK4	62	IPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHV
76	US6	62	VPTTTIRRHVDLLVGAAtFCSAMYVGDLCGSVFLiSQLFTFSPRQHETVQDCNCSIYPGHV
			######################################
68	IND8	62	VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHV
			<u> </u>
67	IND5	62	VsTTTIRhHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHV
			<u>, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>
73	SW2	62	VPTTTT1RRHVDLLVGAAAFCSVMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHV
			<u>,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>
63	HK3	62	VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCS1YPGHV
			<u> </u>
66	HK8	62	VPTTTTRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHV
			<u>,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>
61	D3	62	VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQeCNCSIYPGHV
74	тз		<u>,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>
74	T3	62	VPTKTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHV
	*****		
65	HK5	62	VPTTaIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHV
71	S45		<u> </u>
/1	545	62	VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHETVQDCNCSIYPGHV
72	SA10		<u>,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>
12	SAIO	62	VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRyETVQDCNCSIYPGrV
			<u>,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>
69	P10	62	VPTTAIRRHVDLLVGAAAFCSAMYVGDLCGSV1LVSQLFTFSPRRHWTVQDCNCSIYPGHV
60			<u>,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>
60	D1	62	VPTTAIRRHVDLLVGAAAFCSAMYVGDLCGSVFLISQLFT1SPRRHETVQeCNCSIYPGHV
70			<u> </u>
70	S9	62	VPTTLIRRHVDLLVGAAvFCSAMYVGDLCGSVFLISQLFTiSPRRHETVOnCNCSIYPGHV
60-76			
00-70	consensus		vpTttlRrHVDLLVGAAaFCSaMYVGDLCGSVfLvSQLFTfSPRrheTvQdCNCSiYPGhv

FIGURE 2B

SEO ID NO:	Isolate		
75	T10	123	SGHRMAWDMMMNWSPTTALVvSQLLRIPQAVmDMVtGAHWGVLAGLAYYSMAGNWAKVLI
62	DK1	123	SGHRMAWDMMMNWSPTTALV1SOLLRIPOAVvDMVAGAHWGVLAGLAYYSMAGNWAKVLI
•			
64	HK4	123	SGHRMAWDMMMNWSPTAALVVSOLLR1POAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLI\
••			
76	US 6	123	SGHRMAWDMMNWSPTAALVVSOLLRIPOAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLI
68	IND8	123	SGHRMAWDMMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGILAGLAYYSMVGNWAKVLIV
	21.00		
67	IND5	123	SGHRMAWDMMMNWSPTAALVVSOLLRIPOAVVDMVAGAHWGILAGLAYYSMVGNWAKVLI
0,	21.20		
73	SW2	123	SGHRMAWDMMMNWSPTAALVVSOLLRIPOAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLI\
, -			
63	HK3	123	SGHRMAWDMMMNWSPTAALVVSOLLRIPOAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLI
0.5			
66	HK8	123	SGHRMAWDMMMNWSPTLALVVSOLLRIPOAIVDMVAGAHWGVLAGLAYYSMVGNWAKVLI
00			
61	D3	123	TGHRMAWDMMMNWSPTaALVVSOLLRIPOAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLI
74	Т3	123	TGHRMAWDMMMNWSPTTALVVSOLLRIPOAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLI
65	HK5	123	TGHRMAWDMMMNWSPTTALVVSOLLRIPOAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLI
71	S45	123	TGHRMAWDMMNWSPTaALVVSOLLRIPOAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLI
· <del>-</del>			
72	SA10	123	TGHRMAWDMMMNWSPTLALVVSOLLRIPOAIVDMVAGAHWGVLAGLAYYSMVGNWAKVLI
69	P10	123	sGHRMAWDMMMNWSPTaALVVSOLLRIPOAIlDvVAGAHWGVLAGLAYYSMVGNWAKVLIV
60	D1	123	TGHRMAWDMMMNWSPTTALVVSOLLRIPOAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
70	S9	123	TGHRMAWDMMMWSPTTALVVSQLLRIPQAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
60-76 c	onsensus		sGHRMAWDMMMNWSPTaALVvSQLLRiPQAvvDmVaGAHWGvLAGLAYYSMvGNWAKVLI\

SEO ID NO:	<u>Isolate</u>		
75	T10	184	mLLFAGVDG
62	DK1	184	1LLFAGVDG
64	HK4	184	mLLFAGVDG
76	US6	184	lllfagvDG
68	IND8	184	MLLFAGVDG
67	IND5	184	MLLFAGVDG
			111111111
73	SW2	184	MLLFAGVDG
63	нкз	184	MLLFAGVDG
03	111.03	101	
66	HK8	184	MLLFAGVDG
	-		1111111111
61	D3	184	MLLFAGVDG
74	Т3	184	1LLFAGVDG
			11111111
65	HK5	184	MLLFAGVDG
71	S45	184	MLLFAGVDG
72	SA10	184	MLLFAGVDG
69	P10	184	1111111111
0,5	P10	104	IIIIIIIII
60	D1	184	MLLFAGVDG
70	S9	184	MLLFAGVDG
, 0	39	184	FILLERAGVUG
60-76	consensus		$\mathtt{mLLFAGVDG}$

### FIGURE 2C

SEO ID NO:	<u>Isolate</u> T2	1	AQV:nTsrgymvTnDCSneSiTwQLQAAvLHvPgCiPCErlGnTsRCwiPvtPnVavRQPG
78	T4	1	AQVKNTtnSYMVTNDCSNDSITWQLQAAVLHVPGCVPCEktGNTSRCWIPVSPNVAVRQPG
79	Т9	1	AeVKNTSTSYMVTNDCSNDSITWQLQAAVLHVPGCVPCErVGNaSRCWIPVSPNVAVQRPG
80	US10	1	
77-80	consensus		$\verb"aqVkNTstsYMVTNDCSNdSITWQLqAAVLHVPGCvPCE-vGNtSRCWIPVsPNVAVPG"$
SEO ID NO:	Isolate		
77	T2	62	ALTOGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPrrHWFVQeCNCSIYPGTI
78	T4	62	ALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWFVQdCNCSIYPGTI
79	<b>T</b> 9	62	ALTOGLETHIDMYVMSATLCSALYVGDLCGGVMLAAOMFIiSPOHHWFVOECNCSIYPGTI
80	US10	62	ALTQGLRTHIDMVVMSATLCSALYVGDfCGGmMLAAQMFIvSPrHHsFVQECNCSIYPGTI
77-80	consensus		${\tt ALTQGLRTHIDMVVMSATLCSALYVGD1CGGvMLAAQMFIvSP-hHwFvQeCNCSIYPGTI}$
SEO ID NO:	<u>Isolate</u>		
77	T2	123	TGHRMAWDMMMNWSPTATMILAYAMRVPEVIiDIigGAHWGVMFGLAYFSMQGAWAKViVI
78	T4	123	TGHRMAWDMMNWSPTATMILAYAMRVPEVIlDIvSGAHWGVMFGLAYFSMQGAWAKVVVI
79	T9	123	TGHRMAWDMMNWSPTtTMILAYAMRVPEVIIDIISGAHWGVMFGLAYFSMQGAWAKVVVI
80	US10	123	TGHRMAWDMMNWSPTaTlILAYvMRVPEVIIDIISGAHWGVlFGLAYFSMQGAWAKVVVI
77-80	consensus		${\tt TGHRMAWDMMMNWSPTaTmILAYaMRVPEVIiDIisGAHWGVmFGLAYFSMQGAWAKVvVI}$
SEQ ID NO:	Isolate		
77	T2	184	LLLAAGVDA
78	<b>T4</b>	184	LLLAAGVDA
79	Т9	184	LLLtagvda
80	US10	184	 LLLaAGVDA

77-80

consensus

LLLaAGVDA

### FIGURE 2D

SEO ID NO: 82 83 84	DK11 SW3 T8	1	VEVEN'SSSYYATNDCSNnSITWQLTNAVLHLPGCVPCENDNGTLHCWIQVTPNVAVKHRG
81	DK8	1	VEVRNISSSYYATNDCSNNSITWQLTdAVLHLPGCVPCENDNGTLRCWIQVTPNVAVKHRG
81-84	consensus		VEVRN-SssyyatndcsnnsitwQltnavLhLpgcvpcenDngtL-cwiQvTpnvavKhRg
SEO ID NO:	DK11		ALTHNIRAHiDMIVMAATVCSALYVGDvCGAVMIVSQAFIvSPEhhhFTQECNCSIYQGhI
83 84	SW3	-	ALTHNIRAHVDMIVMAATVCSALYVGDmCGAVMIVSQAFIISPERHNFTQECNCSIYQGrI
	Т8		ALTHNLRTHVDVIVMAATVCSALYVGDVCGAVMIaSQAFIISPERHNFTQECNCSIYQGHI
81	DK8	62	ALTHNLRTHVDVIVMAATVCSALYVGDVCGAVMIvSQAliisPERHNFTQECNCSIYQGHI
81-84	consensus		ALTHNLR-HvD-IVMAATVCSALYVGDvCGAVMIvSQAfliSPErHnFTQECNCSIYQGhI
SEO ID NO:	<u>Isolate</u> DK11	123	TGHRMAWDMMLNWSPTLTMILAYAARVPELVLEVVFGGHHGVVFGLAYFSMQGAWAKVIAI
83	SW3	123	TGHRMAWDMMLNWSPTLTMILAYAARVPELVLEVVFGGHWGVVFGLAYFSMQGAWAKVIAI
84	T8	123	TGHRMAWDMMLNWSPTLITMILAYAARVPELVLEVVFGGHWGVVFGLAYFSMQGAWAKVIAI
81	DK8	123	TGHRMAWDMMLNWSPTLTMILAYAARVPELaLqVVFGGHWGVVFGLAYFSMQGAWAKVIAI
81-84	consensus		TGHRMAWDMMLNWSPTLTMILAYAARVPELvLeVVFGGHWGVVFGLAYFSMQGAWAKVIAI
SEO ID NO:	Isolate DK11	184	LLLVAGVDA
83	SW3	184	LLLVAGVDA
84	Т8	184	LLLVAGVDA
81	DK8	184	LLLVAGVDA

81-84

consensus

LLLVAGVDA

### FIGURE 2E

86 87 88 90 89	<u>Isolate</u> DK12 HK10 S2 S54	1 LEWRNVSGLYVLTNDCsNSSIVYEADDVILHTPGCVPCVQDGNTSTCMTSVTPTVAVRY  1 LEWRNVSGLYVLTNDCsNSSIVYEADDVILHTPGCVPCVQDGNTSTCMTSVTPTVAVRY  1 LEWRNTSGLYVLTNDCSNSSIVYEADDVILHTPGCVPCVQDGNTSTCMTPVTPTVAVRY  1 LEWRNTSGLYVLTNDCSNSSIVYEADDVILHTPGCVPCVQDGNTSTCMTPVTPTVAVRY  2 LEWRNTSGLYVLTNDCSNSSIVYEADDVILHTPGCVPCVQDGNTSTCMTPVTPTVAVRY  1 LEWRNTSGLYVLTNDCSNSSIVYEADDVILHTPGCVPCVQDGNTSTCMTPVTPTVAVRY
86-90	consensus	LEWRNLSGLYvLTNDCsnSSIVYEADDVILHTPGCVPCVQDGNTStCWTpVTPTVAVRY
SEO ID NO 86 87 88 90	DK12 HK10 S2 S54	62 ATTASIRSHVDLLVGAATMCSALYVGDvCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPG
89	S52	62 ATTASIRSHVDLLVGAATLCSALYVGDMCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPG
86-90	consensus	ATTASIRSHVDLLVGAATmCSALYVGDmCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPG
SEQ ID NO: 86 87 88 90 89 86-90	DK12 HK10 S2 S54 S52	123 SGHRMAWDMMMNWSPAVGMVVAHVLRLPQTLFDIIAGAHWGImAGLAYYSMQGNWAKVA  113 SGHRMAWDDMMNWSPAVGMVVAHVLRLPQTLFDIIAGAHWGILAGLAYYSMQGNWAKVA  123 SGHRMAWDDMMNSPAVGMVVAHVLRLPQTVETDIIAGAHWGILAGLAYYSMQGNWAKVA  124 SGHRMAWDDMMNSPAVGMVVAHVLRLPQTVETDIIAGAHWGILAGLAYYSMQGNWAKVA  125 SGHRMAWDDMMNSPAVGMVVAHILRLPQTVETDIIAGAHWGILAGLAYYSMQGNWAKVA  126 SGHRMAWDDMMSPAVGMVVAHILRLPQTLFDIIAGAHWGILAGLAYYSMQGNWAKVA  127 SGHRMAWDMMNSPAVGMVVAHILRLPQTLFDIIAGAHWGILAGLAYYSMQGNWAKVA  128 SGHRMAWDMMNSPAVGMVVAHILRLPQTLFDIIAGAHWGILAGLAYYSMQGNWAKVA
86-90	consensus	SGHRMAWDMMNWSPAVGMVVAHvLRLPQT1FDIiAGAHWGI1AGLAYYSMQGNWAKVA
SEQ ID NO: 86 87 88 90	Isolate DK12 HK10 S2 S54	184 MVMFSGVDA          184 MVMFSGVDA           184 MVMFSGVDA          184 MVMFSGVDA
89	S52	 184 MIMFSGVDA
86-90	consensus	MvMFSGVDA

### FIGURE 2F

94 2 93 2 93-94 consensus (26	6 1	VNYhNASGVYHITNDCPNSSImYEAEHHILHLPGCVPCVReGNQSRCWVALTPTVAAPYIG
SEO ID NO:         Isolate           94         Z           93         Z           93-94 consensus (Z6	7 62 6 62	APLESIRRHVDLMVGAATVCSALYIGDLCGGVFLVGQMFSFQPRRHWTTQDCNCSIYAGHV
94 Isolate 93 Z 93-94 consensus (Z6	6 123	TGHRMAWDMMMNWSPTTTL\LAQVMRIPSTLVDLLTGGHWGiLiGVAYFCMQANWAKVILV
SEO ID NO: Isolate 94 Z 93 Z 93-94 CONSENSUS (Z6	6 184	LFLYAGVDA

### FIGURE 2G

SEO ID NO			
98	SA5	1 VPY	RNASGVYHVTNDCPNSSIVYEADNLILHAPGCVPCVkegNVSRCWVQITPTLSAPNIG
100	SA7	1 VPY	RNASGVYHVINDCPNSSIVYEADNLILHAPGCVPCVRONNVSRCWVQITPTLSAPNLG
97	SA4	1 VPY	RNASGVYHVTNDCPNSSIVYEADNLILHAPGCVPCVRODNVSkCWVQITPTLSAPNLG
96	SA1	1 VPY	RNASGVYHVTNDCPNSSIVYEADsLILHAPGCVPCVRQDNVSRCWVQITPTLSAPtfG
99	SA6		RNASGVYHVINDCPNSSIVYEADDLILHAPGCVPCVRkDNVSRCWVhITPTLSAPSLG
101	SA13		RNASGVYHVTNDCPNSSIVYEADDLILHAPGCVPCVRqgNVSRCWVqITPTLSAPSLG
96-101	consensus	VPY	RNASGVYHVTNDCPNSSIVYEADnLILHAPGCVPCVrqdNVSrCwVqITPTLSAPnlG
SEO ID NO	: Isolate		
98	SA5	62 AVT	APLRRvVDYLAGGAALCSALYVGDACGAVFLVGQMFtYRPRQHTTVQDCNCSIYSGHI
100	SA7	62 AVT	APLRRAVDYLAGGAALCSALYVGDACGAVFLVGQMFsYRPRQHTTVQDCNCSIYSGHI
97	SA4		APLRRAVDYLAGGAALCSALYVGDACGAVFLVGQMFTYRPRQHTTVQDCNCSIYSGHI
96	SA1	62 AVTI	APLRRAVDYLAGGAALCSALYVGDACGAVFLVGQMFTYRPRQHTTVQDCNCSIYSGHI
99	SA6		APLRRAVDYLAGGAALCSALYVGDvCGAlFLVGQMFTYRPRQHaTVQDCNCSIYSGHI
101	SA13		APLRRAVDYLAGGAALCSALYVGDaCGAvFLVGQMFTYsPRrHnvVQDCNCSIYSGHI
96-101	consensus	AVT	APLRRaVDYLAGGAALCSALYVGDaCGAvFLVGQMFtYrPRqHttVQDCNCSIYSGHI
SEO ID NO:	Isolate		
98	SAS	123 TGH	RMAWDMMMNWSPTTALVMAQvLRIPQVVIDIIAGGHWGVLFAvAYFASAANWAKVVLV
100	SA7	123 TGH	RMAWDMMNWSPTTALVMAQLLRIPQVVIDIIAGGHWGVLFAAAYFASAANWAKVVLV
97	SA4	123 TGH	MAWDMMNWSPTTALLMAOLLRIPOVVIDIIAGGHWGVLFAAAYFASAANWAKViLV
96	SA1		RMAWDMMNWSPTTALLMAQMLRIPQVVIDIIAGGHWGVLFAAAYFASAANWAKVVLV
99	SA6	123 TGH	MAWDMMNWSPatalyMaQMLRIPQVVIDIIAGGHWGVLFAAAYFASAANWAKVVLV
101	SA13	123 TGH	RMAWDMMMNWSPtTALVMAQ1LRIPQVVIDIIAGAHWGVLFAAAYYASAANWAKVVLV
96-101	consensus	TGH	RMAWDMMMNWSPtTALvMAQ1LRIPQVVIDIIAGgHWGVLFAaAYfASAANWAKVvLV

FIGURE 2G

SEO ID NO:	<u>Isolate</u>		
98	SA5	184	LFLFAGVDg
100	SA7	184	LFLFAGVDA
			111111111
97	SA4	184	LFLFAGVDA
96	SA1	184	LFLFAGVDg
99	SA6	184	LFLFAGVDA
99	SAb	184	IIIIIIIII
101	SA13	184	LFLFAGVDA
96-101	conceneus		T.PT.PACUDA

### FIGURE 2H

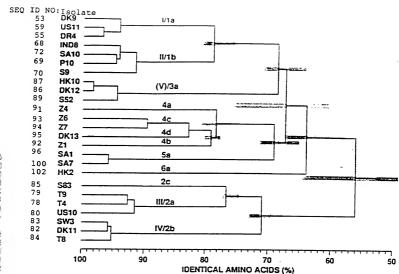
81-84 85 77-80 86-90 60-76 52-59 91 93-94 95 92 96-101 102	O: Genotype (IV/2b) (2c) (III/2a) (V/3a) (II//b) (I/1a) (4c) (4d) (4d) (4b) (5a) (6a)	1 1 1 1 1 1 1 1	VEVRNISASI VEVKNISASI AQVANTSESI AQVANTSESI YEVENVSGYI YQVRNSEGLI YNYENSSGVI YNYENSSGVI VHYENASGVI VHYENASGVI VPYENASGVI VPYENASGVI VPYENASGVI LTYGNSSGLI	MPT VLT NVT NVT HVT HVT HVT HVT HVT	INDCSN INDCSN INDCSN INDCSN INDCPN	SSIV SSIV SSIV SSIV SSIV SSIV SSIV SSIV	WQLEC WQLQ! YEADI YEALI YEALI YETDY YETE! YEADI	AVLH AVLH OWILH OMINH OMILH HILH HILH HILH HILH MILLH LLILH	TTPGC TTPGC TTPGC ISPGC ILPGC ILPGC ILPGC ILPGC ILPGC	VPCERT VPCERV VPCVQE VPCVRE VPCVMT 1 PCVRV VPCVRE VPCVRT VPCVRT	ANVSR GNtSt CNTSt CNTST GNAST GNTSR GNQSR GNKST ENTSR dNVST	CWVP CWIP CWVA CWVA CWVP CWVA CWVS CWVP	VAPNLAI VSPNVAV VTPTVAV LEPTLAA VEPTVAT VTPTVAV LTPTVAV LTPTVAA LTPTVAA ITPTLSA	SQP( QQP( RYVC RNas RDG! AHP( SYIC QHLI PYPI Pn1(
52-102	consensus		Y	т	NDC N	s		н	PGC	PC		CW	P	
SEO ID NO 81-84 81-84 77-80 86-90 60-76 52-59 91 93-94 95 92 96-101 102	D: Genotype (IV/2b) (2c) (V/3a) (V/3a) (II/1a) (I/1a) (4c) (4d) (4d) (5a) (6a) Consensus	62 62 62 62 62 62 62 62 62 62 62	ALTHNLRCHV ALTKGLRAHI ALTYGLRTHI ATTASIRSHV VPTTEIRFHV LPRHV LPATQLRRHI APLESIRRHV APLESIRRHV APLESMRRHV APLESMRRHV APLERHRHV TPATGFRRHV	DILL DLL DLL DLL DLM DLM DLM DYL	VMSAT VMSAT VGAAT VGAAT VGAAT VGAAT VGAAT VGAAT AGAAV	VCSA LCSA FCSA LCSA LCSA VCSA LCSA MCSA LCSA VCSS	LYVGD LYVGD LYVGD MYVGD LYVGD LYVGD LYVGD LYIGD FYIGD LYVGD	VCGA 1 CGG mCGA DLCGS DLCGS DLCGG DLCGG DLCGG DLCGG DLCGG DLCGG	LMLAI VFLVO VFLVO VFLVO AFLMO VFLVO VFLVO VFLVO	AQVVVV AQMFIV SQAFTF SQLFTf SQMITF SQMFSF SQLFTF SQLFDF SQMFLY	SPQHH SPQHH RPRRH SPRTH RPRRH QPRRH QPRRH RPRRH TPRGH	TFVQE WFVQe QTVQO WTTQO WTTQI WTTQI WTTQI WTTQI ELVQI WTVQI	ECNCSIY CCNCSIY CCNCSIY CCNCSIY CCNCSIY CCNCSIY CCNCSIY CCNCSIY CCNCSIY CCNCSIY	PGRI PGTI PGH1 PGhV PGHI TGHI AGHI TGHI PGHV SGHI TGHV
SEO ID NO	: Genotype													
81-84	(IV/2b)	123	TGHRMAWDMM	LNW	SPTLT	A.I.TN	YAARV	PELV	Levv	сснис	MEGT.	AVECL	MCAWAY	
85	(2c)	123	TGHRMAWDMM	MNW	SPTTT	MLLA.	YLVRT	PRVT	LDTV	CCHWC	VMECT	DVFCN	CCCMVA	UTUT
77-80	(III/2a)	123 '	TGHRMAWDMM	MNW.	SPTaT	nILA	Yamrv	PEVI	iDTis	CAHWG	VmFGI.	AVECM	IOCD MARK	VIVI
86-90	(V/3a)	123	SGHRMAWDMM	MNW.	SPAVG	AVV	HVLRL	POT1	FDI i	GAHWG	Ilagi	AYYSN	OCNWAK	UATi
60-76	(II/1b)	123	sGHRMAWDMM	MNW.	SPTaAl	LVvs	OLLRi	POAv	vDmVa	GAHWG	vI.AGI.	AVVSN	LONWAK	VT.TV
52-59	(I/1a)	123 '	TGHRMAWDMM	MNW:	SPTLA	LVVA	OLLRi	POAi	LDMI	GAHWG	VIAGI	AYESM	UCNWAK	VIVV
91	(4a)	123 '	TGHRMAWDMM	MNW:	SPTTTI	LLLA	DIMRV	PTAF	LDMV	AGGHWG	VLAGL	AYESM	OGNWAK	יי.דעיע
93-94	(4c)	123 '	TGHRMAWDMM	MNW.	SPTTTI	LlLA	OVMRI	PSTL	VDLLa	GGHWG	vLvGl	AYFEM	OANWAK	V.TT.V
95	(4d)	123 1	TGHRMAWDMM	MNW:	SPTATI	LVLA	OLMRI	PGAM	VDLL	AGGHWG	ILVGI	AYESM	OANWAK	VIIV
92	(4b)	123	SGHRMAWDMM	MNW:	SPTSAI	LIMA	DILRI	PSIL	GDLLT	GGHWG	VLAGL	AFFSM	OSNWAK	UTTU
96-101	(5a)	123 1	TGHRMAWDMM	MNW:	SPtTAI	LVMA	Ollri	POVV	IDIL	<b>LGaHWG</b>	VLFAal	AYFAS	AANWAK	Uvt.V
102	(6a)	123	TGHRMAWDMM	MNW:	SPTTTI	LVLS	SILRV	PEIC	ASVII	GGHWG	ILLAV	AYFGM	AGNWLK	VLAV
52-102	consenus		GHRMAWDMM	NW	SP		R	P		G HWG	,	A.	wĸ	v

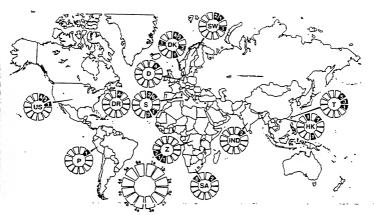
SEO ID NO:	Genotype		
81-84	(IV/2b)	184	LLLVAGVDA
85	(2c)	184	LLLTAGVEA
77-80	(III/2a)	184	LLLaAGVDA
86-90	(V/3a)	184	MVMFSGVDA
60-76	(II/lb)	184	mLLFAGVDG
52-59	(I/la)	184	LLLFaGVDA
91	(4a)	184	LFLFAGVDA
93-94	(4c)	184	LFLfAGVDA
95	(4d)	184	LFLFAGVDA
92	(4b)	184	LFLFAGVEG
96-101	(5a)	184	LFLFAGVDa
102	(6a)	184	LFLFAGVEA
52-102	consensus		GV

r	4
•	v
F	=
:	ť
5	٠
٠	•
1	Ł

f1v	MAN	MIA	GUMLA	AVFLV AVFLV AVF	SOURCE OF STREET	AVFLV AVFLV AVFLV AVFLV AVFLV BIFLV
30 3GBV	AVMI - M - AVMI - V - AVMI	4	GWILL GWILL GWILL GWILL		THE LICENSE AND	111111
30 de	> E>>	>	ЦС	PEEEE		444478
alYv	7277	AL-V V AIMLA	2242	44444 44444	0.00   0.00	1 2 2 2 2 2 2 2
200 210 220 230 240 250 260 270 280 250 260 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 280 280 280 280 280 280 280 280 28	IOVT-NVAVKHRGALTHNL-AH: MIVMA-TVAL-V- IOVT-NVAVKHRGALTHNL-TH-MIVMA-TVAL-V- IOVT-NVAVKHRGALTHNL-THV-VIVMA-TVAL-V- IOVT-NVAVKHRGALTHNL-THV-VIVMA-TVAL-V-		:::::	EEEEE		-Wegwyse, voir Trishricavribe, Rev Trido ALL. V. A - Wydonyse, voir Trishricavribe, Rav Trido ALL. V. A - Wydonyse, voir Trishricavribe, Rav Trido ALL. V. A - Wydonyse, voir Trishricavribe, Rav Trido ALL. V. A - Wronger, voir Trishricavribe, Rav Trido ALL. V A - Wronger, voir Trishricavribe, Rav Trido ALL. V A - Wronger, - AlTrial Prastpatgf, RNV Trido - Well - KI - I
/gaA	FEEF	SE.	SWS	49999	WORSE VOORS VOOR VOOR VOOR VOOR VOOR VOOR VOO	8 000000 a
D11,	EESS	-	2222	33333	35555555555555555555555555555555555555	
260 Rrhv	翻翻	E		E E E E E		OIT-TISABIGATIAPE RAY ILAGG AL OIT-TISABIGATIAPE RAY ILAGG AL OIT-TISABIGATIAPE RAY ILAGG AL OIT-TISABEGATIAPE RAY ILAGG AL OIT-TISABEGATIAPE RAY ILAGG AL OIT-TISABEGATIAPE RAY ILAGG AL RAY-TLAIPABETBAL GRAY ILAGG AL
tt.		g		TAST	TITITITITITITITITITITITITITITITITITITI	TAPPE
50 agap	GOOGG		E SO	GAT	PASY OF THE PROPERTY OF THE PR	as treas
arne	PASS.	9	2008 2008 2588	WRY	AAARA AARAN	SAPE SAPE SAPE
Ptva	NAVA AVA	Ž	NA SA			effere f
240 vavt		ErTanvSRvPVa-NlAisOPGALTKGL-aHI-11VMS-TV	- EKTGNTSR - I PVS - NVAV COPGALIQGL-THI - MVVMS - TI - EKVGNTSR - I PVS - NVAVORPGALITGGL-THI - MVVMS - TI - ERIGNTSR - I PVS - NVAV RPGALITGGL-THI - MVVMS - TI - ERIGNTSR - I PVT - NVAV RAPGALITGGL-THI - MVVMS - TI	VOOGSTST - TSVT - TVARYVORATIASI - SHV-LLVOR-TM VOOGSTST - TSVT - TVARXVORATIASI - SHV-LLVOR-TM VOOGSTST - TPVT - TVARXVORATIASI - SHV-LLVOR-TM VOOGSTST - TPVT - TVARXVORATIASI - SHV-LLVOR-TL - VOOGSTSS - TPVT - TVARXVORATIASI - SHV-LLVOR-TL	S. S. IVERLADDIN T. V. VIRENISSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. S. S. VIRENADDIN T. V. VIRENISSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. S. VIRENADDIN T. V. VIRENISSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. S. VIRENADDIN T. V. VIRENISSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. S. VIRENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIRENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIRENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIRENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIN T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIL T. V. VIRENASSE NALT TLARBINSCROPTTI. BHY LLUDA AND S. VIVENADDIL T. V. VIRENASSE NALT TRARBING CONTROL BHI LLUDAS TRABENT CONTROL BHI LLUDAS TRABENT CONTROL BHI LLUDAS TRABENT CONTROL BHI LLUNGS TRABENT CONTROL BENEVAL BHI LLUNGS TRABENT CONTROL BHI LLUNGS TRABENT CONTR	00000 a
*.		-N	*****	THE STATE	T T S S S S S S S S S S S S S S S S S S	V VkegNVSR VO V VRODNVSR VO V VRODNVSR VO V VRADNVSR - VO V VRADNVSR - VD V VRAGNVSR - VD
gr	NGTI	aNv	GRAN	E E E E E E	O C C C C C C C C C C C C C C C C C C C	O COO COO COO COO COO COO COO COO COO C
230 Cvre	VENDNGTLH- VENDNGTLH- VENDNGTLR- VENDNGTLR-	딤	ERVE	55555	<u> </u>	NESES E
*Ç	>>>>	-	>>>-	>>>>	>>> =>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	>>>>>
0 HtPC	LTNAVL-LV LTNAVL-LV LTGAVL-LV LTNAVL-LV	÷	>>>>	HHHHH		
22 dail	LTNAVI LTNAVI LTNAVI LTNAVI	<b>GAVI</b>	AAAAA AAGGA		DOMINIO DE LA COMPANIO DEL COMPANIO DE LA COMPANIO DEL COMPANIO DE LA COMPANIO DEL COMP	ADNILII ADNILII ADNILII ADDILII ADAMII
/eaa		NOLe	ITWOLGAAVL-V- ITWOLGAAVL-V- ITWOLGAAVL-V- ITWOLGAAVL-V-		INVERDAMIN VANTERAMIN	IVYEADNLI IVYEADNLI IVYEADNLI IVYEADDLI IVYEADDLI IVIEADAMI
110 Siv	FEFF	-s-IvWOLegAVL-t	EEEE	-S-IVYEADDVIL -S-IVYEADDVIL -S-IVYEADDVIL -S-IVYEADDVIL -S-IVYEADDVIL	National	S-IVYEADNLIL-A -S-IVYEADNLIL-A -S-IVYEADNLIL-A -S-IVYEADBLIL-A -S-IVYEADBLIL-A -S-IVYEADBLIL-A -S-IVYEADBLIL-A -S-IVYEADBLIL-A
CBNB	AS -n ITWOL	-S-	လဲလဲလဲ	o ppioioi	N	
Ð	أذذذ	:		ن ن ن ن	ERDANSSOL HV  FREEDRINGSOL HV  FREEDRING	>>>>>
200 7 7	2222	S.E	SEE E	77777		20000000000000000000000000000000000000
nvag	NISS NISS NTS	dTg	NTS NTS NTS	NVS NTS NTS NTS NTS	REVENUESCH HV RE	VPYRNASGV-HV VPYRNASGV-HV VPYRNASGV-HV VPYRNASGV-HV VPYRNASGV-HV VPYRNASGV-HV VPYRNASGV-HV
/evr	VEVRNISSS-YAS VEVRNISSS-YAS VEVRNISSS-YAS VEVRNISSS-YAS	VEVKdTqdS-Mp-	aOVKNTtnS-MV- vOVKNTSTS-MV- AeVKNTSTS-MV- AGVRNTSTG-MV-	LEWRNVSGL-VL LEWRNVSGL-VL LEWRNTSGL-VL LEWRNTSGL-1L LEWRNTSGL-VL	NO OCCUPANTAL DE LA CONTRACTOR DE LA CON	VPYRNASGV-HV- VPYRNASGV-HV- VPYRNASGV-HV- VPYRNASGV-HV- VPYRNASGV-HV- VPYRNASGV-HV- ILYGNSSGI-HI-
Isolate	8311	3	T4 US10 T9 T2	DK12 HK10 S2 S54 S52	IND8 SW425 S	SAS SA1 SA1 SA13 HK2
IBO	SW3 DK8	583	T2 T2	SSS	NI	20000000
ξ. 2.						
SEQ ID NO: 52-102	82 83 84	85	80 77 77	88888 9089 9089	80000000000000000000000000000000000000	1029991
Genotype	Q		28	38	g   _	
Geno	IV/2b	2c	111/2a	(V)/3a	11/11 11/18	S S B

	380 Gvda	- VVFGL-YFSMGGA-A-IAILLLVA-DA VVFGL-YFSMGGA-A-IAILLLVA-DA - VVFGL-YFSMGGA-A-IAILLLVA-DA - VVFGL-YFSMGGA-A-IAILLLVA-DA	eA	- VMFGL YFSWGGA-A - VVILLLAA- DA - VIFGL YFSWGGA - VVILLLAA- DA - VMFGL YFSWGGA - YVILLLAA- DA - VMFGL YFSWGA - IVILLLAA- DA	M. NORWIANTI, L. CTIPDITA A. IMAGI. YSROGN: A. ALIMWRS. DA M. MORWIANTI, L. CTIPDITA A. ILIGIL YSROGN: A. ALIMWRS. DA M. NORWIANTI, L. CTIPDITA A. ILIGIL YSROGN: A. ALIMWRS. DA M. MORWIANTI, L. CTIPDITA A. ILIGIL YSROGN: A. ALIMWRS. M. AVGWIANTIL L. CTIPDITA A. ILIGIL YSROGN: A. ALIMWRS.	222	200	200		385	200	288	3	SAS.	44	A.A VIAGI - YESMVGN-ALIVLLERBDA A.A VIAGI - YESMVGN-ALIVLLERAPA VIAGI - VESMACN-AII VIII I BA	VLAGL-YFSMOGN-AVLVLFLFALA	TITILILAOVM-I-STIVDLLA-GVLVGL-YFSMQAN-AILVLFLFADA-TITILILAOVM-I-STIVDLLH-GILIGO-YFCMQAN-A-ILVLFLVADA	ILVGi-YFSMQAN-AILVLFLFADA	VLaG1-fFSMOsN-AILVLFLFAeG	- A	VLFAA-YFASAAN-AYLVUELFFADA VLFAA-YFASAAN-AVLVUELFFADA VLFAA-YFASAAN-AVLVUELFFADA	- DA	5
	.11fa		LLL	iiiii Iiiii	TWAFS	ALLED ALLED							2		LLEF		FLF	LFLF!	LFLF!	FLF			FE	1
	370 kvliv	ANT T	VMFGL-YFSMOGB-AIVILLLTAeA		-Ailin			- E		TI.	127				33	33:	· vIV	:: :::	- IIV	-ILV		333		TTELVISBIL-V-ELCASVII-GILLAV-ILGUMGN-IIAVLELEAEN
	nwar	4444	8-A-	4444	AAAAA	A-A-	Á	A-A	ZZZ	ZZ	A A	AAA	2	AA	d d	AAA	N-A-	N.A.	N-A-	N-A-	44	444 555	A P	1
	) Esmvg	20000	FSMOG		NOOOO NA SWAOO	YSMVC	SMV	YSMVC	N SWA	YSMV	YSMV	YSMAC	TOTAL STATE	FSMV	FSMVC	FSMVC	FSMOC	FSMO2	FSMO	FSMO	FASA	FASA	VASA	T dille
	360 glAyf	2000 2000 2000 2000 2000 2000 2000 200	GL-Y	7777 5555	55555	999	122	GL-Y	999	300	155	1225	1	125	150	555	GL-Y	GL-Y GV-Y	Gi-Y	G1 - £	AA-Y	*** \$ <b>\$</b> \$	AA-Y	- A
	Gvla	XXXX	VMF	ANY TANA	SEEES	111		33			55	555			55		7	VI.	- II	VL	33	355	7	
	350 aGaHi			4444	4444	44.	444	AA	444	144	A	444		444	ė d	44	g S	-6-	a.G	t-G-	99	999	A-a-	-6-1
	1dmv	LEVV	VIGI			VMOV					V DWC				G.	ŽĮ.	FLOM	TICK CAN	-qamVDLLa-G	IdDLI	ZIDII		/IDI	1000
	340 Pgav	V-ELV V-ELV V-ELA	i - EVI	V-EVI V-EVI V-EVI	55555	i-oay	100	00			NO.		5	500	Š.	66 60 HH:	4-7	I-ST	I-gar	I-Bi	\$5 1 1 1		ð	4-e1
	qllR	XAA-	YIV.	YAM-		-i-i-	2000	111			XX.		3	300			N N	MAC	NO1M-	Oir-	00 11-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-		OIL	3811-
	330 talwa	HILL	TMILE	HILL	CONTRACTOR OF THE PROPERTY OF	MAN A		33				ZAZ			VIV	NA NA NA NA NA NA NA NA NA NA NA NA NA N	TITLLLAOIM-V-tAflDMVA	TLLL	-TaTLVLAOIM-I	TBALiMAQiL-I-BilgDLLt-G-	ALVE		ALVM.	E LV 1.
	SPtt	LTLTMILAYAA-V-ELVLEVVF-G- -LTLTMILAYAA-V-ELVLEVVF-G- -LTLTMILAYAA-V-ELALQVVF-G- LTLTMILAYAA-V-ELVL@VVF-G-	TETMILAYIV-i-EVILDIVE-G-	TATMILAYAM V EVILDIVS-A- TATHILAYYM V EVILDIIS-A- TETMILAYAM V EVILDIIS-A- TATMILAYAM V EVILDIIG-A-	- AV	E E	i i	A.	ĖĖ	446	E.E	EE	ŀ	E.	Ę	- TTALVVAQLL I QALDMIA A - TTALVVAQLL I QALDMIA A				TB	##	TTALLMAOLL-I-OVVIDIIA-G-  TTALLMAOML-I-OVVIDIIA-G- 		
9	1 TATULINI T	4444	W	EEEE	×××××	ΣĖ	EΣΣ	×Σ	ĖĖ	ÉÉ	EΣΣ	××	Ė	EΣΣ	××	:ÉÉ:	E	Σ×	Σ	W	EΕ	×××	Σ	Ε
PIGURE	290 300 300 310 320 320 330 340 350 360 370 380 390 300 300 300 370 300 300 300 300 300 30		:																		:::			
_	itGHB		IT			SSS	200	55	H	SSS	ם בי	- P-HVS		TIH-d-1	E			H	-	BA	H		i.	5
	310 YpGhi	1.0-hir- 1.0-rir- 1.0-Hir- 1.0-Hir-	-P-rIT	P-TIT P-TIT-G-	1. P-HLS- -1. P-HLS- -1. P-HLS- -1. P-HLS-	44	444	44	44	44		44	H			HH	- I	I-A-HIT	I-t-HiT	BvH-d-	H-S-1	SEE	S	-I-t-HvI
	A*		1 : 1					: :	::				:										-11	:
	300 tvQd	S-AFIVS ENHIPT ESAFIIS ERHNFT ESAFIIS ERHNFT ESAFIIS ERHNFT ESAFIIS ERHNFT ESAFIIS ERHNFT ES	A- vvvVS - OHHL FV - E	A-MFIVS-OHHWFV-d- A-MFIVS-IHHBFV-E- A-MFIIS-GHHWFV-E- A-MFIVS-RRHWFV-E	G-AFTER-RRHOTV-T- G-AFTER-RRHOTV-T- G-AFTER-RRHOTV-T- G-AFTER-RRHOTV-T- G-AFTER-RRHOTV-T-	22	200	25	22	201 221	22	S-LFTFS-RRHETA-D-	12-6		Ė	G-LFTFS-RRHWIT-D- G-LFTFS-RRIWIT-D-	G-LFT1S-RRHWIT-D	G-MFSFO-RRHWTT-D	G-LFFFO-BRHWIT-D-	G-LFGFR-RRHWIT-D	20	G-MFTYR-ROHTTV-D-G-MFTYR-ROHTTV-D-G-MFTYR-ROHTTV-D-G-MFTYR-ROHTTV-D-G-MFTYR-ROHTTV-D-G-MFTY-D-G-MFT-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFTY-D-G-MFT-D-G-MFTY-D-G-MFTY-D-G-MFT-D-G-MFT-D-G-MFT-D-G-MFT-D-G-MFT-D-G-MFT-D-G-MFT-D	VV.	G-lFTfg-RRHwtV-D-
	Prrhw	ERHN ERHN ERHN	-OHHC	器器	XXXXX	RRHE	RRHE	REHE	RRHE	RGH	RRY	E E	- RRH	RRH	REH		REH	RRH	E E	- RRH	- ROH	200	ž	- RRH
	290 ftf8	STITE	VVVS	FIVS FIVS FITS FITS	TTTTT RANKER	FTFS	TES	TES	TES	TES	TES	TISS	FTIS	TIPS STITES	TES	FFS	177	FSFO	5	FGFR	FEYR		FTYB	Ttd
		9000	A.	AAAA	99999	SS	, oo	S	ww.	o co	, coc	ooo.	S	990	ÖÖ	9 6		00	9	9	00	000	00	5
	Isolate	I		0.	29	82	o.ma		.0.		.S.			<u>. =</u> .					-				2	~
		DK11 SW3 DK8	583	T4 US10 T2	DK12 HK10 S2 S54 S52	EE	HK	24	EH	OSC	S	SS	DI	SSC	Š	SWI	24	256	DK1	Z1	SAS	SA1	SAS	H
	NO:																							
	SEQ ID NO: 52-102	8833 841 841	85	8.86 6.00 7.00 8.00 7.00	90000	68	633	22.5	65	496	91-1 701	1920	9	non Non	un. ∪ 4.	വവാ	57	63	4	92	100	766 766	101	102
	Genotype	2b		111/2a	V) /3a				d d							ns								
	Genc	IV/2b	20	H.	3				11/11				1		I	1/	9	5	43	4		58		68





# FIGURE 6A

SEO ID NO: ISOLATE	
108 DR4	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
108 DK4	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
103 DK7	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCCAACCGTCGCCCACAGG
104 0311 105 S14	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
106 SW1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
107 S18	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCCAACCGTCGCCCACAGG
107 516	1 Alabana Alabana and Alabana
103-108 consensus	ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
SEO ID NO: ISOLATE	
108 DR4	62 ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
103 DK7	62 ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
104 US11	62 ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
105 S14	62 ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCCAGGGG
106 SW1	62 ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
107 S18	62 ACGTLAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
20.	
103-108 consensus	ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
SEO ID NO: ISOLATE	
108 DR4	123 CCCTAGATTGGGTGTGCGCGCGACGAGGGAAGACTTCCGAGCGGTCGCAACCTCGAGGTAGA
103 DK7	123 CCCTAGATTGGGTGTGCGCGCGCGAGGGAAGACTTCCGAGCGGTCGCAACCTCGAGGTAGA
104 US11	123 CCCTAGATTGGGTGTGCGCGCGAGGAGGAGGACTTCCGAGCGTCGCAACCTCGAGGTAGA
105 S14	123 CCCTAGATTGGGTGTGCGCGCGAGGAGGAGGACTTCCGAGCGTCGCAACCTCGAGGTAGA
106 SW1	123 CCCTAGATTGGGTGTGCGCGCGACGACGAAGACTTCCGAGCGGTCGCAACCTCGAGGTAGA
107 S18	123 CCCTAGATTGGGTGTGCGCGCGACGACGAAGACTTCCGAGCGGTCGCAACCTCGCGGTAGA
107 310	113 CCCTACATIOGOTOTOCOCOCCACCAACCTCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCAACCTCCCAACCTCCCAACCTCCCAACCTCCCCAACCTCCCCAACCTCCCCAACCTCCCAACCTCCCAACCTCCCAACCTCCCAACCTCCCAACCTCCCAACCTCCCAACCTCCCCAACCTCCAACCTCCAACCTCCAACCTCCAACCTCCAACCTCCAACCTCCAACCTCCAACCTCAACCTCAACCTCAACCTCCAACCTCAACAA
103-108 consensus	CCCTAGATTGGGTGTGCGCGCGaCGAGGAAGACTTCCGAGCGGTCGCAACCTCGaGGTAGA
SEO ID NO: ISOLATE	184 CGTCAGCCTATCCCCAAGGCCCCCTCGACCGCCACGACCGAC
108 DR4	184 CGTCAGCCTATCCCCAAGGCGCGTCGGCCCGAGGCAGGACCTGGGCTCAGCCCGGGTACC
108 DR4 103 DK7	184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC
108 DR4 103 DK7 104 US11	184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC
108 DR4 103 DK7 104 US11 105 S14	184 CGTCAGCCTATCCCCAAGGCÃCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGAGGACCTGGGCTCAGCCCGGGTATC
108 DR4 103 DK7 104 US11 105 S14 106 SW1	184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC
108 DR4 103 DK7 104 US11 105 S14	184 CGTCAGCCTATCCCCAAGGCÃCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGAGGACCTGGGCTCAGCCCGGGTATC
108 DR4 103 DK7 104 US11 105 S14 106 SW1	184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC
108 DR4 103 DK7 104 US11 105 S14 106 SW1 107 S18	184 CSTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGGTACC 184 CSTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CSTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCACGTTGGCCCGAGGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCGCTTGGCCCGAGGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCGCTTGGCCCGAGGGGCAGGACCTGGGCTCAGCCCGGGTACC
108 DR4 103 DK7 104 US11 105 S14 106 SW1 107 S18 103-108 CONSENSUS	184 CSTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGGTACC 184 CSTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 CSTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCACGTTGGCCCGAGGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCGCTTGGCCCGAGGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCGCTTGGCCCGAGGGGCAGGACCTGGGCTCAGCCCGGGTACC
108 DR4 103 DK7 104 US11 105 S14 106 SN1 107 S18 103-108 CONSENSUS  SEO ID NO: ISOLATE	184 CSTCAGCCTATCCCCAAGGCĂCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGGTACC 184 CSTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGCTAGGCCGGGGTACC 184 CSTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCGCTGGGCCCGAGGGGCAGGACCTGGGCTAGCCCGGGTATC 184 CSTCAGCCTATCCCCAAGGCCTGGGCCCGAGGGCAGGACCTGGGCTAGCCCGGGTACC CGTCAGCCTATCCCCAAGGC-CGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC
108 DR4 103 DK7 104 US11 105 S14 106 SW1 107 S18 103-108 CONSENSUS  SEO ID NO: ISOLATE 108 DR4	184 GOTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGCTAGCCCGGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGC-CGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCCAAGGC-CGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCCCCCCGTGG
108 DR4 103 DK7 104 US11 105 S14 106 SN1 107 S18 103-108 CONSENSUS  SEO ID NO: ISOLATE 108 DR4 103 DR7	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGACGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 COTCAGCCTATCCCCAAGGCGCTGGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCGCTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGC-CGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCCCCCCTGG
108 DR4 103 DK7 104 US11 105 S14 106 SW1 107 S18 103-108 CONSENSUS  SEO ID NO: ISOLATE 108 DR4 103 DK7 104 US11	184 GOTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 GGTCAGCCTATCCCCAAGGCCGGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTTCTCCCCCTTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTTCTCCCCCTTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTTGCGGTTGGGCGGGATGGCTCCTTCTCCCCCTTGG 247 CTTGGCCCCTCTATGGCAATGAGGGCTTGCGGGTGGGCGGATTGGCTCCTTCTCCCCCTTGG 247 CTTGGCCCCTCTATGGCAATGAGGGCTTGCGGGTTGGGCGGGATTGGCTCCTTCTCCCCCTTGG
108 DR4 103 DK7 104 US11 105 S14 106 SN1 107 S18 103-108 CONSENSUS  SEO ID NO: ISOLATE 108 DR4 103 DK7 104 US11 105 S14	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 COTCAGCCTATCCCCAAGGCGCTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGGC-CGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGG-CGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTTCTCCCCCTGG
108 DR4 103 DK7 104 US11 105 S14 106 SN1 107 S18 103-108 CONSENSUS  SEO ID NO: ISOLATE 108 DR4 103 DK7 104 US11 105 S14	184 GOTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 GGTCAGCCTATCCCCAAGGCCGGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTTCTCCCCCTTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTTCTCCCCCTTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTTGCGGTTGGGCGGGATGGCTCCTTCTCCCCCTTGG 247 CTTGGCCCCTCTATGGCAATGAGGGCTTGCGGGTGGGCGGATTGGCTCCTTCTCCCCCTTGG 247 CTTGGCCCCTCTATGGCAATGAGGGCTTGCGGGTTGGGCGGGATTGGCTCCTTCTCCCCCTTGG
108	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 185 CGTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCCCGAGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGATTGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGATTGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGGGGATTGGGCGATTGCTCCCCCTTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGGC 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGGC 246 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGGATTGGCTCCTCCCCCCTGGCCCCTCTGCCCCCCTGGCCCCTCTGCCCCCC
108   DR4     103   DK7     104   US11     105   S14     106   SW1     107   S18     103-108   CONSENSUS     SEO ID NO:   ISOLATE     108   DR4     103   DK7     104   US11     105   S14     106   SW1	184 GOTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 GGTCAGCCTATCCCCAAGGCGCGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCGCTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCTCCCCCCTGG
108	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 185 CGTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCCCGAGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGATTGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGATTGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGGGGATTGGGCGATTGCTCCCCCTTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGGC 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGGC 246 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGGATTGGCTCCTCCCCCCTGGCCCCTCTGCCCCCCTGGCCCCTCTGCCCCCC
108 DR4 103 DK7 104 US11 105 S14 106 SW1 107 S18 103-108 CONSENSUS  SEO ID NO: ISOLATE 108 DR4 103 DK7 104 US11 105 S14 106 SW1 107 S18 103-108 CONSENSUS	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 185 CGTCAGCCTATCCCCAAGGCGCTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCCCGAGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTSGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGATTGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGATTGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGGGGATTGGGCGATTGCTCCCCCTTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGGC 245 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGATTGGCTCCTTCCCCCCTGGC 246 CTTGGCCCCCTTATGGCAATGAGGGCTTGCGGGTTGGGCGGGATTGGCTCCTCCCCCCTGGCCCCTCTGCCCCCCTGGCCCCTCTGCCCCCC
108	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 185 CGTCAGCCTATCCCCAAGGCGCTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCCCGAGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATGAGGGTGCGGGTGGGCGGGATGGCTCCTGTCCCCCTGG 245 CTTGGCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 247 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGGATGGCTCCTGTCCCCCCTGG 247 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGGATGGCTCCTGTCCCCCCTGG 248 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGGATGGCTCCTGTCCCCCCTGG 247 CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGGATGGCTCCTGTCCCCCCTGG CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGGATGGCTCCTGTCCCCCCTGGC CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGGATGGCTCCTGTCCCCCCTGGC CTTGGCCCCCTTATGGCAATGAGGGCTTCGGGTTGGGCGGGATGGCTCCTGTCCCCCCTGGCCCCCCTGGCCCCCTCTATGGCAATGAGGGCTTCGGGTTGGGGGGGG
DR4   DR4	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGAACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 COTCAGCCTATCCCCAAGGCGCTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  CGTCAGCCTATCCCCAAGGCCCGAGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTGGGCGGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCTCCCCCGTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCTCCCCCCTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCCCTCTATGGCAATGAGGGCTGCGGTTGGGGGGGATGGCTCCTGTCCCCCCCTGG 246 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCCCCCCCTGG 247 CTTGGCCCCCCTCTATGGCAATGAGGGCTGCGGTTGGGGGGATGGCTCCTGTCCCCCCCTGG 248 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCCTGG 247 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGGGATGGCTCCTGTCCCCCCCTGG 248 CTTCGCCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGGGATGGCTCCTGTCCCCCCCTGGC 248 CTCTCGGCCCCTCTATGGCAATGAGGGCTCCGGGTTGGGCGGAATGGCTCCTGTCCCCCCCTGGC 247 CTCTCGCCCCTCTATGGCAATGAGGGCTCCGGGTTGGGCGGAATGGCTCCTGTCCCCCCCTGGCCCCCCCTGGCGCGGATGGCTCCTGTCCCCCCCTGGCCCCCCTGGCGCGGATGGCTCCTGTCCCCCCCTGGCGCGGATGGCTCCTGTCCCCCCCTGGCCCCCCGCGCGCG
108	184 GOTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTCGGCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCGCTGGCCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC  CGTCAGCCTATCCCCAAGGCCCGAGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  245 CTTGGCCCCTCTATGGCAATAAGGGCTCGGGTGGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGGATGGCTCCTGTCTCCCCCGTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGGATGGCTCCTGTCTCCCCCTTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTTGG 246 CTCTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTTGG 247 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGGGATGGCTCCTGTCCCCCCTTGG 248 CTCTCGGCCCTAGCTGGGGCCCCACAGACCCCCGGCGAAGGTCGGCCAAATTTTGGGTAAAGTT
108	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 COTCAGCCTATCCCCAAGGCGCTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  CGTCAGCCTATCCCCAAGGCCCGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCCTTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGAATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGAATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCCTCTATGGCAATGAGGGCTGCGGGATGGCTCCTGTCCCCCCCGTGG 245 CTTGGCCCCCCTCTATGGCAATGAGGGCTCCGGGTTGGGGGATGGCTCCTGTCCCCCCCGTGG 245 CTTCGGCCCTAGCTGGGGCCCAAGAACCCCCGGCGAAGGCCCCAAATTTGGGTAAAGGT 306 CTCTCGGCCCAAGGCCCCAAGAACCCCCGGCGAAGGTCGCGAATTTGGGTAAAGTC 306 CTCTCGGCCTAGCTGGGGCCCAAGAACCCCCGGCGAAGGCCCCAATTTGGGTAAAGTC
108	184 GOTCAGCCTATCCCCAAGGCĂCGTCGGCCCAAGGCAGAACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGCCCAAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGCCCAAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTCAGCCCGAGGGCAGGACCTGGGTCAGCCCGGGTATC 184 CGTCAGCCTATCCCCAAGGCCGTGGCCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCCCGAGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCCCGAGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATAAGGGCTGCGGTSGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTSGGCGGATGGCTCCTGTCTCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGATGGCTCCTGTCTCCCCCTGGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGATGGCTCCTGTCCCCCCTGGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCCCCCCTGGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGAATGGCTCCTGTCCCCCCTGGG 246 CTCTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGAATGTTGGGTAAGGT CTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGCAAGTTGGGCAAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGCAAGTTGGGCAAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGCAAGTGGCGCAATTTTGGGTAAAGTC 306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCCAATTTGGGTAAAGGT
108	184 GOTCAGCCTATCCCCAAGGCĂCGTCGGCCCAAGGCAGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGCCCAAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTGGCCCAAGGGCAGGACCTGGGTCAGCCCGGGTACC 184 GGTCAGCCTATCCCCAAGGCACGTCAGCCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC 185 CGTCAGCCTATCCCCAAGGCCGTCGGCCCGAGGGCAGGACCTGGGTCAGCCCGGGTACC CGTCAGCCTATCCCCAAGGCCCGAGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 245 CTTGGCCCCTCTATGGCAATAAGGGCTCGGGTGGGCGGATGGCTCAGCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGCTTCGGGTTGGGCGGGATGGCTCCTGTCTCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGATGGCTCCTGTCCCCCCTGG 246 CTCTGGCCCCTCTATGGCAATGAGGGCTGCGGTTGGGCGGAATGCTCCTGTCCCCCCTGG 247 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGAATTTGGGTAAGGT 248 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGCGCAAGTTGGGCAATTTTGGTAAAGTC 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGCGAAGGTCGCCCAATTTGGGTAAAGTC 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGCGAAGGTCGCCAATTTGGGTAAAGTC 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGCGTAGGTCGCCAATTTGGGTAAAGTC
108	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 COTCAGCCTATCCCCAAGGCGCTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  CGTCAGCCTATCCCCAAGGC-CGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGTTGCGGGTTGGGCGGATTGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATTGGCTCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATTGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATTGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATTGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGAATGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGAATGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGAATTTGGGTAAGGT 306 CTCTCGGCCCACGAATGAGGCCCCACGAACCCCCGGCGAAGGTCGCGCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGGCGAAGGTCGCGCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGGCGTAGGTCGCGCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGGCGTAGGTCGCGCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCCGGCGTAGGTCGCCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGCGTAGGTCGCCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGCGTAGGTCGCCAATTTGGGTAAGGT
108	184 COTCAGCCTATCCCCAAGGCĂCGTCGGGCCGAGGGCAGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC 184 COTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC 184 COTCAGCCTATCCCCAAGGCGCTGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  CGTCAGCCTATCCCCAAGGC-CGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGATGGCTCCTGTCCCCCCGTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTTGGGCGGATGGCTCCTGTCTCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGTTGCGGGTTGGGCGGATTGGCTCCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATTGGCTCTGTCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATTGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATTGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGATTGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGAATGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGAATGGCTCCTGTCCCCCCCTGG 245 CTTGGCCCCTCTATGGCAATGAGGGCTGCGGATGGGCGGAATTTGGGTAAGGT 306 CTCTCGGCCCACGAATGAGGCCCCACGAACCCCCGGCGAAGGTCGCGCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGGCGAAGGTCGCGCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGGCGTAGGTCGCGCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGGCGTAGGTCGCGCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCCGGCGTAGGTCGCCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGCGTAGGTCGCCAATTTGGGTAAGGT 306 CTCTCGGCCTAGCTGGGGCCCCACGAACCCCCGGCGTAGGTCGCCAATTTGGGTAAGGT

### FIGURE 6A

SEO ID NO: ISOLATE	
	367 ATCGACACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATcCCGCTCGTCGGCGCCC
	367 ATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCC
104 US11	367 ATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCC
	367 ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCC
	367 ATCGATACCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
107 S18	367 ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
103-108 consensus	ATCGALACCCTcACGTGCGGCTTCGCCGACCTCATGGGGTACATaCCGCTCGTCGGCGCCC
105 100 Combanded	
SEO ID NO: ISOLATE	
108 DR4	428 CcCTTGGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGAGTTCTGGAAGACGGCGTGAA
	128 CTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGCGTGAA
	428 CTCTCGGAGGCGCTGCCAGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
	428 CcCTCGGgGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
106 SW1	428 CTCTEGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
107 S18	4 2 8 CTCT_GGAGGCGCTGCCAGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
107	· · · · · · · · · · · · · · · · · · ·
103-108 consensus	CtCT-GGaGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGgGTTCTGGAAGACGGCGTGAA
SEO ID NO: ISOLATE	
	489 CTATGCAACAGGGAALCTTCCTGGTTGCTCTTTCTCTATCTTCCTTTTGGCLLTGCTCTCT
103 DK7	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTTTGGCCCTGCTCTCT
104 US11	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTC
	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTcCTaGCCCTGCTTTCT
	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTC
107 S18	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTC
103-108 consensus	CTATGCAACAGGGAAcCTTCCTGGTTGCTCTTTCTCTATCTTCCTtcTqGCccTGCTcTCT
202 200 20110011000	
SEO ID NO: ISOLATE	
	550 TGCtTGACCGTGCCCGCaTCGGCC
103 DK7	550 TGCCTGACCGTGCCCGCTTCGGCC
104 US11	550 TGCCTGACTGTGCCCGCTTCAGCC
	550 TGCCTGACTGTGCCCGCTTCAGCC
	550 TGCCTGACaGTGCCCGCGTCAGCC
107 S18	550 TGECTGACEGTGCCCGCGTCAGCE
103-108 consensus	TGccTGACtGTGCCCGCtTCaGCc

	110000 00
SEO ID NO: ISOLATE	
	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
118 IND8	
111 D1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
112 US6	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
113 P10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
114 DK1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
115 T10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
116 SW2	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
122 HK4	1 ATGAGCACGAATCCTAAACCTCAAAGAAAGACCAAACGTAACACCAACCGCCGCCCACAGG
109 SA10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
110 S45	1 ATGAGCACGAATCCTAAACCTCAAAGACAAACCAAACGTAACACCAACCGCCGCCCACAGG
123 P8	1 ATGAGCACGACTCCTAAACCTCAAAGAAAAACCAAACGTAACACCAgCCGCCGCCCACAGG
124 T3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
120 HK3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
121 HK5	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
109-124 consensus	ATGAGCACGAaTCCTAAACCTCAAAGAaAAACCAAACGTAACACCAaCCGCCGCCCACAGG
103-124 COMBEMBUB	ATOMOGRATICE TANACET CHANGE AND
SEO ID NO: ISOLATE	
119 S9	62 ACGTLAAGTTCCCGGGCGGTGGLCAGATCGTCGGTGGAGTTTACCTGTTGCCGCGCAGGGG
117 IND3	62 ACGTCAAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
118 IND8	62 ACGTCAAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
111 D1	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
112 US6	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
113 P10	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
114 DK1	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
115 T10	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGGTTTACCTGTTGCCGCGCAGGGG
116 SW2	62 ACGTCAAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
122 HK4	62 ACGTLAAGTTCCCGGGCGGTGGCCAGATCGTCGGTGGAGTTTACCTGTTGCCGCGCAGGGG
109 SA10	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTCTALCTGTTGCCGCGCAGGGG
110 S45	62 ACGTCAAGTTCCCGGGLGGCGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
123 P8	62 ACGTTAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
124 T3	62 ACGTTAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
120 HK3	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
121 HK5	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGGTTTACCTGTTGCCGCGCAGGGG
109-124 consensus	ACGT-CAAGTTCCCGGG-CGGLGGLCAGATCGTLGGTGGAGTLTA-CTGTTGCCGCGCAGGGG
SEO ID NO: ISOLATE	
119 S9	123 CCCCAGGTTGGGTGTGCGCGCAACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
117 IND3	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
118 IND8	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
111 D1	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
112 US6	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
113 P10	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
114 DK1	123 CCCCAGGTTGGGTGTGCGCGGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
115 T10	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
116 SW2	123 CCCC-GGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
122 HK4	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
109 SA10	123 CCCCAGGTTGGGTGTGCGCGCGACGAGGAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
110 S45	123 CCCCAGGTTGGGTGTGCGCGGGCGACTAGGAAGACTTCCGAGCGGTCaCAACCTCGTGGAcGG
123 P8	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGATCGCAACCTCGTGGCAGG
124 T3	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
120 HK3	123 CCCCAGGTTGGGTGTGCGCGCGACCAGGAAGACTTCAGAGCGGTCGCAACCTCGTGGAAGG
121 HK5	123 CCCCAGGTTGGGTGTGCGCGGGCCAGGAAGACTTCcGAGCGGTCGCAACCTCGTGGAAGG
109-124 consensus	CCCCaGGTTGGGTGTGCGCGCgACtAGGAAGACTTCcGAGCGGTCgCAACCTCGTGGaaGG
SEO ID NO: ISOLATE	
119 S9	184 CGACAACCTATCCCCAAGGCTCGCCatCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC

84892\_1

117	IND3	184 CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCCTGGGCTCAGCCCGGGTACC
118	IND8	184 CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGTAGGGCCTGGGCTCAGCCCGGGCACC
111	D1	184 CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCCTGGGCTCAGCCCGGGTACC
112	US6	184 CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC
113	P10	184 CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC
	DK1	184 CGACAACCTATCCCCAAGGCTCGCCGGGCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC
114		
115	T10	184 CGACAGCCTATCCCCAAGGCTCGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC
116	SW2	184 CGACAACCTATCCCCAAGGCTCGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCLGGGTACC
122	HK4	184 CGACAACCTATCCCCAAGGCTCGCCAACCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC
109	SA10	184 CGACAACCTATCCCCAAGGCTCGCCAGCCCGAGGGCAGGACCTGGGCCCAGCCCGGGTACC
110	S45	184 CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGCAGGGCCTGGGCCCAGCCCGGGCALC
123	P8	184 CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCCTGGGCTCAGCCCGGGCACC
124	T3	184 CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCCTGGGCTCAGCCCGGGTACC
120	HK3	184 CGACAACCTATCCCCAAGGCTCGCC&ACCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC
121	HK5	184 CGACAACCTATCCCCAAGGCTCGCCGACCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC
109-124	consensus	CGACAaCCTATCCCCAAGGCTCGCCggCCCGAGGGCAGGgCCTGGGCCCAGCCcGGGtAcC
SEO ID NO		
119	59	245 CTTGGCCCCTCTAcGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTCACCCCGtGG
117	IND3	245 CTTGGCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTCACCCCGCGG
118	IND8	245 CTTGGCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTCACCCCGCGG
111	D1	245 CTTGGCCCCTCTATGGCAACGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTCACCCCGCGG
112	US6	245 CTTGGCCCCTCTATGGCAACGAGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGG
113	P10	245 CTTGGCCCCTCTATGGCAATGAGGGCLTGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGG
114	DK1	245 CTTGGCCCCTCTATGGCAATGAGGGCATGGGTGGGCAGGATGGCTCCTGTCACCCCGGGG
115	T10	245 CTTGGCCCCTCTATGGCAATGAGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGLGG
116	SW2	245 Cottggcccctttatggcaatgaggcatgggtatgggtaggatggctcctgtcccccccgcgg
122	HK4	245 CTTGGCCCCTCTATGGCAATGAGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGCGG
109	SA10	245 CTTGGCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGG
110	S45	245 CTTGGCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTCACCCCGTGG
123	245 P8	
		245 CTTGGCCCCTCTATGcCAATGAGGGCTTGGGGTGGGCTGGGATGGCTCCTGTCACCCCGCGG
124	T3	245 CTTGGCCCCTCTATGGCGACGAGGGCATGGGTGGGCAGGATGGCTCCTGTCACCCCGCGG
120	HK3	245 CTTGGCCCCTCTATGGCAACGAGGGCATGGGGTGGGCAGGATGGCTCCTGTCACCCCGCGG
121	HKS	245 CTTGGCCCCTCTATGGCAALGAGGGCATGGGTGGGCAGGATGGCTCCTGTCACCCCaLGG
109-124	consensus	CtTGGCCCCTCTAtGgCaAtGAGGGC-TGGGgTGGGCaGGATGGCTCCTGTCaCCCCgcGG
SEO ID NO		
119	S9	306 cTCTCGGCCTAGTTGGGGCCCCAatGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
117	IND3	306 tTCTCGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTC
118	IND8	306 CTCTCGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
111	D1	306 CTCCCGGCCTAGTTGGGGCCCCACcGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
112	US6	306 CTCCCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
113	P10	306 CTCTCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
114	DK1	306 CTCTCGGCCTAGTTGGGGCCCCAacGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
115	T10	306 CTCcCGGCCTAGTTGGGGCCCCACaGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
116	SW2	306 CTCTCGGCCTAGTTGGGGCCCCACLGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
122	HK4	306 CTCTCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
109	SA10	306 CTCTCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGLAATTTGGGTAAGGTC
110	S45	306 CTCCCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
123	P8	306 CTCCCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
124	T3	306 CTCCCGGCCTAATTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGLAATcTGGGTAAGGTC
120	HK3	306 CTCTCGGCCTAATTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
121	HK5	306 CTCTCGGCCTAgTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGtAATTTGGGTAAGGTC
109-124	consensus	cTCtCGGCCTAgTTGGGGCCCCAcgGACCCCCGGCGTAGGTCGCGtAATtTGGGTAAgGTC
SEO ID NO	: ISOLATE	
119	59	367 ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
117	IND3	367 ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
118	IND8	367 ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
		300010000000000000000000000000000000000
24877 1		

111	D1	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
112	US6	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
113	P10		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
114	DK1		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
115	T10		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
116	SW2		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
122	HK4		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
109	SA10		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
110	S45		ATCGATACCCTCACqTGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
123	P8		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGGCC
124	T3		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCtC
120	HK3		ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGtGCCC
121	HKS		ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGcGCCC
109-124	consensus		ATCGATACCCTCACaTGCGGCTTcGCCGACCTCATGGGGTACATtCCGCTCGTCGGcGccC
SEO ID NO			
119	S9	428	CCCTAGGGGCGCTGCCAGGGCtCTGGCGCATGGCGTCCGGGTtCTGGAGGACGGCGTGAA
117	IND3		CCCTAGGGGGGCCTGCCAGGGCCCTGGCGCATGCGGTCCTGGAGGACGGCGTGAA
118	IND8		CCCTAGGGGGTGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCCTGGAGGACGGCGTGAA
111	D1	428	CCCTAGGGGGTGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
112	US6	428	CCCTAGGGGGCCTGCCAGGGCCLTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
113	P10		CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCTCCGGGTTCTGGAGGACGGCGTGAA
114	DK1	428	CCCTAGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
115	T10		CCCTAGGGGGGCTGCCAGGGCLCTGGCACATGGLGTCCGGGTTCTGGAGGACGGCGTGAA
116	SW2		CCCTAGGGGGCGCTGCCAGGGCCCTGGCgCATGGCGTCCGGGTCCTGGAGGACGGCGTGAA
122	HK4		CCTTAGGGGGGGTTGCCAGAGCCCTGGCACATGGtGTCCGGGTTGTGGAGGACGGCGTGAA
109	SA10	428	CETTAGGGGGCGCTGCCAGgGCCTTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
110	S45		CCCTAGGGGGCGCTGCCAGaGCCTTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
123	P8		CCCTAGGGGGCGTTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTgTGGAGGACGGCGTGAA
124	Т3	428	CCLTAGGGGGCGTTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
120	HK3	428	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA
120 121	HK3 HK5	428	
121	HK5	428	CCCTAGGGGGGTTGCCAGAGCC+TGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGCGTTGCCAGAGCC+TGGCACA-GGTGTCCGGGTTCTGGAGGACGGCGTGAA
		428	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA
121	HK5	428	CCCTAGGGGGGTTGCCAGAGCC+TGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGCGTTGCCAGAGCC+TGGCACA-GGTGTCCGGGTTCTGGAGGACGGCGTGAA
121 109-124	HK5 consensus	428	CCCTAGGGGGGTTGCCAGAGCC+TGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGCGTTGCCAGAGCC+TGGCACA-GGTGTCCGGGTTCTGGAGGACGGCGTGAA
121 109-124 SEO ID NO	HK5 consensus : ISOLATE	428 428	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGCGTTGCCAGAGCCCTGGCACACGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGCGTGCCAGGGCCCTGGCGCALGGCGTCCGGGTLCTGGAGGACGGCGTGAA
121 109-124 SEO ID NO 119	HK5 consensus : ISOLATE S9	428 428 489	CCTASGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCCAGAGCCCTGGCACAGGGTGTCCGGGTTCTTGGAGGACGGCGTGAA CccTAGGGGGGGTGCCAGGGCcCTGGCGCALGGGGTCCGGGTTCTGGAGGACGGCGTGAA CTATGCAACAGGGAACCTCCCCGGTTGCTCTTTCTCTATCTTCTTTGGGTTTGCTTTCCT
121 109-124 <u>SEO ID NO</u> 119 117	HKS consensus : ISOLATE S9 IND3	428 428 489 489	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCCAGGGCCCTGGCACACGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGCGTGCCAGGGCCCTGGCGCALGGCGTCCGGGTLCTGGAGGACGGCGTGAA CTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTGTCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTAGCTTTGCTATCC
121 109-124 SEO ID NO 119 117 118	HK5 consensus : ISOLATE S9 IND3 IND8	428 428 489 489 489	CCTASGGGGGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCCAGAGCCCTGGCACAGGGTGTCCGGGTTCTGGAGGACGGCGTGAA CccTAGGGGGGGTGCCAGGGCCCGGGTGCCCGGGTCCTGGGTTCTGAGGAGACGGCGTGAA CTATGCAACAGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGCTTTTGCTTTCTATCTTCCTTTAGCTTTTGCTTTCTATCTTCCTTTTGGCTTTGCTTTCTCTTTTGCTTTTGCTTTTCTTTTCTTTTCGCTTTTGCTTTTCTTTTCTTTTCGCTTTTGCTTTTCTTTTCTTTTCTTTTTCGTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTT
121 109-124 SEO ID NO 119 117 118 111	HK5 consensus : ISOLATE S9 IND3 IND3 IND8 D1	428 428 489 489 489	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGGTGCCAGGGCCCTGGCACACGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGCGTGCCAGGGCCCTGGCGCACGGGTCCGGGTTCTGGAGGACGGCGTGAA CTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGTATCT CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGTTTCTCTTCTTTTCTTTTTTTT
121 109-124 SEO ID NO 119 117 118 111 112	HKS consensus : ISOLATE S9 IND3 IND8 D1 US6	428 428 489 489 489 489	CCTASGGGGGTTGCCAGAGCCLTGGCACATGCTCTCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCAAGGCCCTGGCCACAGGTTCTGGAGGACGGCGTGAA CcCTAGGGGGGGTGCCAGGGCCCTGGCCACAGGCCCTGGCTCCTGGGTTCTCTGAGGAGACGGCGTGAA CTATGCAACAGGAACTTCCCCGGTTGCTCTTTCTCTATCTTCCTTTAGGTTTTGCTTCTCTCTATCTTCCTTTAGGTTTTGCTTCTCTTTAGCACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTTCTTCTCTTTCCTTTTGGCTTTTCTTC
121 109-124 <u>SEO ID NO</u> 119 117 118 111 112 113	HK5 consensus : ISOLATE S9 IND3 IND3 IND8 D1	428 428 489 489 489 489 489	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGGTGCCAGGGCCCTGGCACACGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGCGCGCAGGCCCTGGCGCACGGGTCCGGGTTCTGGAGGACGGCGTGAA CTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTTTCCTATCTACACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCCTCTATGCACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGTATCTCCTATGCACAGGGAACTTGCCGGTTGCTCTTTCTCTATCTTCCTTTGGCTTTTGCTTCTCTCTTCTCTTTCTT
121 109-124 SEO ID NO 119 117 118 111 112	HKS consensus : ISOLATE S9 IND3 IND8 D1 US6 P10	428 428 489 489 489 489 489	CCTASGGGGGTTGCCAGAGCCLTGGCACATGCTGCTGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCCAGGCCCTGGCACAGGTTCTCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCCAGGGCCCTGGCCACAGGGTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGCTTTTGCTTTCATCTTCCTTTAGCTTTTCTTTC
121 109-124 SEO ID NO 119 117 118 111 112 113 114 115	HKS consensus  : ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10	428 428 489 489 489 489 489 489	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGGTGCCAGGGCCCTGGCACACGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGCGCGCAGGCCCTGGCGCACGGGTCCGGGTTCTGGAGGACGGCGTGAA CTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTGGCTTTGCTGTCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTTATCTTCCTTTTGGCTTTTGCTATC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTTGCTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTGGCTTTGTTCTC CTAGGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTTATCTTCCTCTTTGGCTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTTATCTTCCTCTTTGGCTTTGTTCTTCTCTTTGCTTTTGCTTTTCCTTTTTGCTTTTTT
121 109-124 <u>SEO ID NO</u> 119 117 118 111 112 113 114	HKS consensus : ISOLATE S9 IND3 IND8 D1 US6 P10 DK1	428 428 489 489 489 489 489 489	CCTASGGGGGTTGCAGAGCCLTGGCACATGCTCTCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCAGAGCCCTGGCACAGGGTTCTGGAGGACGGCGTGAA CCTAGGGGGGGTGCAGGGCCTGGCGCACAGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGGTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTCTATCT CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGTTCT CTAGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTTGTTTCT CTAGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTGTTT
121 109-124 SEO ID NO 119 117 118 111 112 113 114 115 116	HK5 consensus : ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10 SW2	428 428 489 489 489 489 489 489 489	CCTAGGGGGCTTGCAGAGCCLTGGCACATGCTGCTGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCAGGCCTGGCACAGGGTTCTGGAGGACGGCGTGAA CcTAGGGGGGGGGCTGCAGGCCTGGCGCACAGGGTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTTGGCTTTGCTATC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTATCTTCCTTTTTGGCTTTTGCTATC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTATCTTCCTTTTGGCTTTTGCTATC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTTATCTTCCTTTTGGCTTTTGCTTTC CTATGCAACAGGGAATTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTTATCTTCCTCTTTGGCTTTTGCTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTTATCTTCCTTTTGGCTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTTATCTTCCTTTTGGCTTTGCTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTTTATCTTCCTCTTTGGCTTTTGCTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTTTATCTTCCTCTTTGGCTTTTGCTTTCC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTATCTTCCTTTTGGCTTTTGCTTTTCC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTTATCTTCCTTTTGGCTTTTGCTTTTCCTTTTTTTT
121 109-124 SEO ID NO 119 117 118 111 112 113 114 115 116 122	HK5 CONSENSUS  ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4	428 428 489 489 489 489 489 489 489	CCTASGGGGGTTGCAGAGCCLTGGCACATGCTCTCGGGTTCTGGAGGACGGCTGAA CCCTAGGGGGGTTGCAGAGCCCTGGCACAGGTTCTGGAGGACGGCTGAA CCTAGGGGGGGTGCAGGGCCTGGCCACAGGTTCTGAGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTTTTCTCTATCTTCCTTTAGGTTTTGCTTTCATCT CTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTCTTATC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTTGCTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGCTTTTGTTCT CTAGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGCTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTATCTTCCTCTTGGCTTTGGTTTGCTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTATCTTCCTCTTGGCTTTGGTTTGCTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTTTGGCTTTGGCTTTGCTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCTCTTGGCTTTGGCTTTGTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCTCTTTGGCTTTGGCTTTGTCTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCTCTTTGGCTTTGGCTTTGTCTCTCTTTTCCTTTTTT
121 109-124 SEO ID NO 119 117 118 111 112 113 114 115 116 122 109	HK5 CONSENSUS  I ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4 SA10	428 428 489 489 489 489 489 489 489 489 489	CCTAGGGGGGTTGCAGAGCCLTGGCACATGCTCTCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCAGAGCCCTGGCACAGGGTTCTGGAGGACGGCGTGAA CCTAGGGGGGGTGCAGGGCCTGGCGCACAGGGTTCTGGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTTTTCTCTATCTTCCTTTAGGTTTTGCATCT CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGCTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTTGTTCT CTAGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTGTTT
121 109-124 SEO ID NO 119 117 118 111 112 113 114 115 116 122 109	HKS CONSENSUS  ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4 SA10 S45	428 428 489 489 489 489 489 489 489 489 489	CCTAGGGGGGTTGCAGAGCCLTGGCACATGCTCTCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCAGAGCCCTGGCACAGGGTTCTGGAGGACGGCGTGAA CCTAGGGGGGGTGCAGGCCTGGCCACAGGGTTCTGAGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTTTTCTCTATCTTCCTTTAGGTTTTGCTTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTTGGCTTTTGTTCT CTATGCAACAGGGAATTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGTTTTGCTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGTTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTTGCTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTTGGTTTTTTTT
121 109-124 SEO ID NO 119 117 118 111 112 113 114 115 116 122 109 110 123	HKS CONSENSUS  ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T100 SW2 HK4 SA10 S45 P8	428 428 489 489 489 489 489 489 489 489 489	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCCAGGCCCTGGCACAGGTGTCCGGGTTCTGGAGGACGGCGTGAA CccTAGGGGGGGGGCCAGGGCCTGGCCACAGGGTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGCTTTTGCTTTCCTATCTTCCTTTAGCTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTAGCTTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTGTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGCTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGCTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTATCTTCCTCTTTGGCTTTGCTTTCTCTTTCCTTTTTCCTTTTTT
121 109-124 SEQ ID NO 119 117 118 111 112 113 114 115 122 109 110 123 124	HKS CONSENSUS  ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3	428 428 489 489 489 489 489 489 489 489 489 48	CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCCAGGCCCTGGCACAGGTGTCCGGGTTCTGGAGGACGGCGTGAA CccTAGGGGGGGGGCCAGGGCCTGGCCACAGGGTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGCTTTTGCTTTCCTATCTTCCTTTAGCTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTAGCTTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTGTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGCTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGCTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTATCTTCCTCTTTGGCTTTGCTTTCTCTTTCCTTTTTCCTTTTTT
121 109-124 SEO ID NO 119 117 118 111 112 113 114 115 116 116 122 109 110 110 123 124 123 124 125 127 127 128 129 129 120 121 121 121 121 122 123 124 125 126 127 127 128 129 129 129 129 129 129 129 129	HKS CONSENSUS  SP IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK5	428 428 489 489 489 489 489 489 489 489 489 48	CCTAGGGGGGTTGCAGAGCCLTGGCACATGCTCTCGGGTTCTGGAGGACGGCTGAA CCCTAGGGGGGTTGCAGAGCCCTGGCACAGGGTTCTGGAGGACGGCGTGAA CCTAGGGGGGGTGCAGGCCTGGCCACAGGGTTCTGAGGTCCGGGTTCTGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTTTTCTCTATCTTCCTTTAGGTTTTGCTTTCC CTATGCAACAGGGAACTTGCCGGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTTGGGTTTTGCTTC CTATGCAACAGGGAATTGCCCGGTTGCTCTTTCTTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCTCTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTGTTT
121 109-124 SEO ID NO 119 117 118 111 112 115 116 122 109 110 123 124	HKS CONSENSUS  ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T110 SW2 HK4 SA10 S45 P8 T3 HK3	428 428 489 489 489 489 489 489 489 489 489 48	CCTAGGGGGGTTGCAGAGCCLTGGCACATGCTCTCGGGTTCTGGAGGACGGCTGAA CCCTAGGGGGGTTGCAGAGCCCTGGCACAGGGTTCTGGAGGACGGCGTGAA CCTAGGGGGGGTGCAGGCCTGGCCACAGGGTTCTGAGGTCCGGGTTCTGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTTTTCTCTATCTTCCTTTAGGTTTTGCTTTCC CTATGCAACAGGGAACTTGCCGGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTTGGGTTTTGCTTC CTATGCAACAGGGAATTGCCCGGTTGCTCTTTCTTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCTCTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTGTTT
121 109-124 SEO ID NO 119 117 118 111 112 113 114 115 116 116 122 109 110 110 123 124 123 124 125 127 127 128 129 129 120 121 121 121 121 122 123 124 125 126 127 127 128 129 129 129 129 129 129 129 129	HKS CONSENSUS  SP IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK5	428 428 489 489 489 489 489 489 489 489 489 48	CCCTAGGGGGGTTGCCAGAGCCLTGGCACATGCTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCCAGAGCCCTGGCACAGGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTGCCAGGGCCCTGGCCACAGGGTTCCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTTTTCTCTATCTTCCTTTAGGTTTTGCTTTCATCT CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGCTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTTGTTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTGTTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGTTTTGTTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTTATCTTTCCTTTTGGTTTTTGTTTTCT CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTTGTTTTC CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTTGGGTTTTGTTTTC CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTTGGGTTTTGTTTTC CTATGCAACAGGGAATTTGCCTGGTTGCTTTTCTTATCTTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTTCCTGGTTTGCTTTTCTTATCTTTCCTTTTTTTT
121 109-124  SEO ID NO 119 117 118 111 112 113 114 115 116 122 109 110 123 124 120 120 121 109-124	HKS CONSENSUS  ISOLATE S9 INDS INDS EP10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK3 CONSENSUS	428 428 489 489 489 489 489 489 489 489 489 48	CCCTAGGGGGGTTGCCAGAGCCLTGGCACATGCTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCCAGAGCCCTGGCACAGGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTGCCAGGGCCCTGGCCACAGGGTTCCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTTTTCTCTATCTTCCTTTAGGTTTTGCTTTCATCT CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGCTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTTGTTTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTGCTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTGTTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGTTTTGTTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTTATCTTTCCTTTTGGTTTTTGTTTTCT CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTTGTTTTC CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTTGGGTTTTGTTTTC CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTTATCTTTCCTTTTTGGGTTTTGTTTTC CTATGCAACAGGGAATTTGCCTGGTTGCTTTTCTTATCTTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTTCCTGGTTTGCTTTTCTTATCTTTCCTTTTTTTT
121 109-124  SEO ID NO 119 117 118 111 112 113 114 115 116 119 119 119 1109 1109 1109 1109 110	HKS CONSENSUS  ISOLATE S9 IND3 IND3 IND8 US6 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK5 CONSENSUS	428 428 489 489 489 489 489 489 489 489 489 48	CCCTAGGGGGGTTGCCAGAGCCLTGGCACATGCTGCTCTGGGGTTCTTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCCAGGCCCTGGCACAGGTTCTCGGGTTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCCAGGGCCCTGGCCACAGGTTCTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTCTATCTCCCTATGCACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTATCT CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTATCT CTATGCAACAGGGAATTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGCTTTGTCTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGCTTTGTCTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTATCTTCCTCTTGGGCTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTATCTTCCTCTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTCTATCTTTCCTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTCCTTTGGGTTTTGTTCTC CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTCCTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTCCTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTCTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCGGTTGCTCTTTCTCTATCTTCCTCTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCTGGTTGCTTTTCTTATCTTCCTCTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCTGGTTGCTCTTTCTTCTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTGCCTGGTTGCTCTTTCTCTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTTCCTTGCTTGTTTTCTTTTTTTT
121 109-124  SEO ID NO 119 117 118 111 112 113 114 115 116 122 109 110 123 124 120 120 121 109-124  SEO ID NO	HKS CONSENSUS  I ISOLATE S9 INDOS IN	428 428 489 489 489 489 489 489 489 489 489 48	CCCTAGGGGGCTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCCAGGCCCTGGCACAGGTGTCCGGGTTCTGGAGGACGGCGTGAA CccTAGGGGGGGGGCAGGCCTGGCGCACAGGGTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACCTCCCGGTTGCTCTTCTCTATCTTCCTTTAGCTTTGCTGTCC CTATGCAACAGGGAACTTCCCCGGTTGCTCTTTCTCTATCTTCCTTTAGCTTTTGCTTTTCCTATCTTCCTTTAGCTTTTGCTTTTCCTTTAGCTTTTGCTTTTCCTTTTGCTTTTTCTTTTCCTTTTTCCTTTTTGCTTTTTCTTTTTCTTTTTCTTTTTT
121 109-124  SEO ID NO 119 117 118 111 112 113 114 115 116 119 119 1100 1100 1121 120 120 121	HKS CONSENSUS  ISOLATE S9 IND3 IND3 IND8 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK5 CONSENSUS  ISOLATE S9 IND3	428 428 489 489 489 489 489 489 489 489 489 48	CCCTAGGGGGGTTGCCAGAGCCLTGGCACATGCTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCCAGGCCCTGGCACAGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTGCCAGGGCCCTGGCCACAGGTTCCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGGTTTGCTTTCC CTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTCTATCT CTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTCTTTGGCTTTTGCTTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGCTTTGTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGCTTTGTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGCTTTGTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTGTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTGGTTTTGCTTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTGTTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTGTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTGTTCTC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCTTTTCTTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCTTTTCTTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCTTTTCTTATCTTCCTTTTGGGTTTTGTTTCT CTATGCAACAGGGAATTTCCCGGTTGCCTTTTCTTATCTTCCTTTTTTTT
121 109-124  SEO ID NO 119 117 118 111 112 113 114 115 116 122 109 110 123 124 120 120 121 109-124  SEO ID NO 119 117 119 117 118	HK5 CONSENSUS  I SOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK5 CONSENSUS  I SOLATE S9 IND3 IND8	428 428 489 4899 4899 4899 4899 4899 489	CCCTAGGGGGCTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCCAGGCCCTGGCACAGGTGTCCGGGTTCTGGAGGACGGCGTGAA CccTAGGGGGGGGGCGCAGGCCCTGGCCACAGGTTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGCTTTGCTTTCCTATCTCCCTTAGCTTTTGCTTTCCTATCTTCCTTTAGCTTTTGCTTTCCTTTAGCTTTTGCTTTTCCTTTAGCTTTTGCTTTTGCTTTTGCACAGGGAACTTCCCCGGTTGCTCTTTTCTCTATCTTCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCCTTTTGGCTTTTTCTTTTCCTTTTTGGCTTTTGCTTTTCCTTTTTGCTTTTTGCTTTTTCCTTTTTT
121 109-124  SEO ID NO 119 117 118 111 112 113 114 115 116 119 110 1123 124 120 120 121 120 121 120 121 121 120 121 121	HKS CONSENSUS  ISOLATE S9 IND3 IND3 IND8 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK5 CONSENSUS ISOLATE S9 IND3 IND3 IND3 IND3 IND3 D1	428 428 489 489 489 489 489 489 489 489 489 48	CCCTAGGGGGGTTGCCAGAGCCLTGGCACATGCTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTTGCCAGGCCCTGGCACAGGTTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGTGCCAGGGCCCTGGCCACAGGTTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGGTTTGCTTTCC CTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTCTATCT CTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGCTTC CTATGCAACAGGGAACTTCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTTGTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTGGGTTTTGTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTTTTGGTTTTGCTTTCC CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTCCTCTTTGGTTTTGCTTCT CTATGCAACAGGGAATTTCCCGGTTGCTCTTTTCTATCTTTCCTTTTGGTTTTTTTT
121 109-124  SEO ID NO 119 117 118 111 112 113 114 115 116 122 109 110 123 124 120 120 121 109-124  SEO ID NO 119 117 119 117 118	HK5 CONSENSUS  I SOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK5 CONSENSUS  I SOLATE S9 IND3 IND8	428 428 489 489 489 489 489 489 489 489 489 48	CCCTAGGGGGCTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA CCCTAGGGGGGGTGCCAGGCCCTGGCACAGGTGTCCGGGTTCTGGAGGACGGCGTGAA CccTAGGGGGGGGGCGCAGGCCCTGGCCACAGGTTCCGGGTTCTGGAGGACGGCGTGAA CCTATGCAACAGGGAACCTCCCGGTTGCTCTTTCTCTATCTTCCTTTAGCTTTGCTTTCCTATCTCCCTTAGCTTTTGCTTTCCTATCTTCCTTTAGCTTTTGCTTTCCTTTAGCTTTTGCTTTTCCTTTAGCTTTTGCTTTTGCTTTTGCACAGGGAACTTCCCCGGTTGCTCTTTTCTCTATCTTCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCTTTTGGCTTTTGCTACTTCCCCTTTTGGCTTTTTCTTTTCCTTTTTGGCTTTTGCTTTTCCTTTTTGCTTTTTGCTTTTTCCTTTTTT

113	P10	550	TGccTGACCATCCCAGCgTCCGCT
114	DK1	550	TGTtTGACCATCCCAGCTTCCGCc
115	T10	550	TGTCTGACCATCCCAGCTTCCGCT
116	SW2	550	TGTCTGACCATCCCAGCTTCCGCT
122	HK4	550	TGTTTGACCATCCCAGCTTCCGCT
109	SA10	550	TGTTTaACCATCCCAGCTTCCGCT
110	S45	550	TGCTTGACCATCCCAGCTTCCGCT
123	P8	550	TGtcTGACCATCCCAGCTTCCGCT
124	T3	550	TGCTTGACCATCCCAGCTTCCGCT
120	HK3	550	TGCTTGACCACCCCAGCTTCCGCT
121	HK5	550	TGtcTGACCACCCCAGtTTCCGCT
100 104			TO LETTER CONTRACTOR C

SEO ID N	O: ISOLATE	
119	S9	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
117	IND3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCCACAGG
118	INDS	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
111	D1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
112	US6	
113	P10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
114	DK1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAACCCAACCGTAACACCGAACCGCCGCCCACAGG
115	T10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
116	SW2	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGG
122	HK4	1 ATGAGCACGAATCCTAAACCTCAAAGAAAGACCAAACGTAACACCAACCGCCGCCACAGG
109	SA10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCCCACAGG
110	S45	1 ATGAGCACGAATCCTAAAACCTAAAGCAAACGTAACACCAACCGCCGCCCACAGG
123	P8	1 ATGAGCACGACTCCTAAACCTCAAAGAAAACCAAACGTAACACCCAGCCGCCCACAGG
124	T3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCCCACAGG
120	HK3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCCCACAGG
121	HK5	1 ATGAGCACGAATCCTAAACCTCAAAGAAAACCAAACGTAACACCAACCGCCGCCCACAGG
108	DR4	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
104	US11	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
105	S14	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
106	SW1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
	S18	1 ATGAGCACBAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
107		
103	DK7	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAGG
103-124	consensus	ATGAGCACGAATCCTAAACCTCAAAGAAAACCTAAACGTAACACCAACCGCCGCCCACAGG
SEO ID N		
119	S9	62 ACGTEAAGTTCCCGGGCGGTGGECAGATCGTCGGTGGAGTTTACCTGTTGCCGCGCAGGGG
117	IND3	62 ACGTCAAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
118	IND8	62 ACGTCAAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
111	D1	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
112	US 6	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
113	P10	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
114	DK1	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
115	T10	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
116	SW2	62 ACGTCAAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
122	HK4	62 ACGTEAAGTTCCCGGGCGGTGGCCAGATCGTCGGTGGAGTTTACCTGTTGCCGCGCAGGGG
109	SA10	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTCTALCTGTTGCCGCGCAGGGG
110	S45	62 ACGTCAAGTTCCCGGGEGGCGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
123	P8	62 ACGTTAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
124	T3	62 ACGTTAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
120	HK3	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
121	HKS	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
108	DR4	62 ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
104	US11	
105	S14	
106	SW1	62 ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
107	S18	62 ACGTLAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
103	DK7	62 ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
103-124	consensus	ACGTcAAGTTCCCGGGcGGLGGLCAGATCGTLGGTGGAGTLTAccTGTTGCCGCGCAGGGG
SEO ID NO		
119	S9	123 CCCCAGGTTGGGTGTGCGCGCAACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
117	IND3	123 CCCCAGGTTGGGTGTGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
118	IND8	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
111	D1	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
112	US6	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
113	P10	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
114	DK1	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
115	T10	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
116	SW2	123 CCCCcGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
122	HK4	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG
	III	115 CCCCCCC11GGACCTCGTGGAAGG

			FIGURE 6C
109 110 123 124 120 121 108 104 105 106 107 103	SA10 S45 P8 T3 HK3 HK5 DR4 US11 S14 SW1 S18 DK7	123 123 123 123 123 123 123 123 123 123	CCCCAGGTTGGGTGTGCGCGGACGAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCACAACCTCGTGGAAGG CCCCAGGTTGGGTGGTGCGCGGACTAGGAAGACTTCCGAGCGGTCACAACCTCGTGGAAGG CCCCAGGTTGGGTGTGCGCGGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGAAGG CCCCAGGTTGGGTTG
103-124	consensus		${\tt CCCcaGgTTGGGTGTGCGCGCgaCtAGGAAGACTTCcGAGCGgTCgCAACCTCGtGGaaGg}$
SEQ ID NO: 119 117 118 111 112 113 114 115 116 122 109 110 123 124 120 123 124 120 121 108 104 105 106	ISOLATE	184 184 184 184 184 184 184 184 184 184	GALAACTATCCCCAAGGTTCGCCGCCGAGGGTAGGGCTTGGGCTCAGCCCGGGTACC GACAACTATCCCCAAGGTTCGCGGCCCGAGGGCAGGCCTTGGGCTCAGCCCGGGTACC GACAACTATCCCCAAGGTTCGCGGCCCGAGGGCAGGGC
103-124	consensus		CGaCAaCCTATCCCCAAGGCtCGcCggcCCGAGGGcAGGgCCTGGGCtCAGCCcGGGtAcC
SEQ ID NO: 119 117 118 111 112 113 114 115 116 122 109 110 123 124 120 121 108 104 105 106	ISOLATE S9 IND3 IND8 D1 US6 P10 DK1 T10 SW2 HK4 SA10 S45 P8 T3 HK3 HK5 DR4 US11 S14 SW1 S18	245 245 245 245 245 245 245 245 245 245	CTTGGCCCCTCTACGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGCAGGATGGCTCCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGCAGGATGGCTCCTGTCACCCCGCGGCTTGGCCCCTTATAGCAACGAGGGCTTGGGGTTGGGGAGGATGGCTCCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAACGAGGGTTGGGGTGGGGAGATGGCTCTGTCACCCCGTGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGAGATGGCTCTGTCACCCCGTGGCTTGGCCCCTCTATAGCAATGAGGGCATGGGGTTGGGAGATGGGTCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGAGATGGGTTCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGAGATGGGTTCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGAGATGGGTTCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGGAGATGGGTTCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGGTTGGGATGGGTTCGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGGTTGGGAGTTGCTCTGTCACCCCGCGGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGGTTGGGAGTTGCTCTTCTCCCCCCGCGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGGTTGGGAGTTGCTCTTCTCCCCCCGCGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGGTTGGGAGTTGCTCTTCACCCCCCGCGCTTGGCCCCTCTATAGCAATGAGGGCTTGGGGTTGGGGTTGGGAGTTGCTCTTCACCCCCCGTGCTTGGCCCCCTTATAGCAATGAGGGCTTGGGGTTGGGGTTGGGAGTAGGCTCTGTCACCCCCGTGCTTTGCCCCCCTTATAGCAATGAGGGTTGGGGTTGGGGTTGGGAGTAGGCTCTGTCACCCCCGTGCTTGCCCCCCTTATAGCAATGAGGGTTGGGGTTGGGGTTGGGAGTAGGCTCCTGTCACCCCCCGTGCTTTGCCCCCCTTATAGCAATGAGGGTTGGGGTTGGGGTTGGCGGGATAGGCTCCTGTCACCCCCCGTGCTTGCCCCCCCTTATAGCAATGAGGGTTGGGGTTGGGGGGGG

	103	DK7	245	$\tt CTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCLCCCCGTGG$
	103-124	consensus		$\tt CtTGGCCCCTCTAtGgCaAtGAGGGCttgGGgTGGGCaGGATGGCTCCTGTCaCCCCgtGG$
	SEO ID NO:	ISOLATE		
	119	59		$\verb"ctctcggcctagttgggccccaatgacccccggcgtaggtcgcgtaatttgggtaaggtc"$
	117	IND3		tTCTCGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAAGTC
	118	IND8		$\tt CTCTCGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC$
	111	D1		$\tt CTCCCGGCCTAGTTGGGGCCCCACcGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC$
	112	US6		CTCCCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
	113	P10		CTCTCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
	114	DK1		CTCTCGGCCTAGTTGGGGCCCCAacGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
	115 116	T10 SW2		CTC-CGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC CTCTCGGCCTAGTTGGGGCCCCACLGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
	122	HK4		CTCTCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGTAATTTGGGTAAGGTC
	109	SA10		CTCTCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
	110	S45		CTCCCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
	123	P8		CTCCCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
	124	T3		CTCCCGGCCTAATTGGGGCCCCACGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
	120	нкз		CTCTCGGCCTAATTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
	121	HK5		CTCTCGGCCTAGTTGGGGCCCCACGGACCCCCGGCGTAGGTCGCGLAATTTGGGTAAGGTC
-	108	DR4		CTCTCGGCCTAGCTGGGGCCCCACaGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
ad No.	104	US11		CTCTCGGCCTAGCTGGGGCCCCACgGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
2	105	S14		CTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
3	106	SW1		CTCTCGGCCTAGCTGGGGCCCTACAGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTC
3	107	S18		CTC-CGGCCTAGCTGGGGCCTACAGACCCCCGGCGTAGGTCGCGCAATTTGGGCAAAGTC
-	103	DK7	306	CTCtCGGCCTAGCTGGGGCCCACAGACCCCCGGCGCAGGTCGCGCAATTTGGGtAAAGTC
4				
è	103-124	consensus		cTCtCGGCCTAgtTGGGGCCCcAc-GACCCCCGGCGtAGGTCGCGtAATtTGGGtAAgGTC
j				
=				
	SEO ID NO:	ISOLATE	260	. MOG. W. COOPO, C. MOG. COMP. CO. C.
3	119 117	S9 IND3		ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
3	118	INDS		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
	111	DI		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCCC
ž	112	US6		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
-	113	P10		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
2	114	DK1		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
1	115	T10		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
	116	SW2		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
	122	HK4		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
	109	SA10		ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
	110	S45	367	ATCGATACCCTCACgTGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
	123	P8	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGgCC
	124	Т3	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCtC
	120	HK3		ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGtGCCC
	121	HK5		ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
	108	DR4 US11	367	ATCGACACCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
	104 105	US11 S14		ATCGATACCCTLACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCC
	105	SW1		ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCC
	105	SWI S18		ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
	103	DK7	367	ATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCCATCGATACCCTLACGTGCGGCTTCGCCGACCTCATGGGGTACATaCCGCTCGTCGGCGCCC
	103-124	consensus		${\tt ATCGALACCCTCACaTGCGGCTTCGCCGACCTCATGGGGTACATLCCGCTCGTCGGcGccC}$
		ISOLATE		
	SEO ID NO-			
	SEO ID NO:		422	
	119	S9	428 428	CCCTAGGGGGCGCTGCCAGGGCCTGGCGCATGGCGTCCGGGTtCTGGAGGACGGCGTGAA
			428	CCCTAGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCCTGGAGGACGCCGTGAA
	119 117	S9 IND3	428 428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCCTGAGGAGGACGGCGTGAA CCCTAGGGGGTGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCCTGAGGACGGCGTGAA
	119 117 118	S9 IND3 IND8	428 428 428	CCCTAGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCCTGGAGGACGCCGTGAA

175

j. A

fil

#### FIGURE 60

```
113
              P10
                      428 CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGCCGTGAA
114
              DK1
                      428 CCCTAGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
                      428 CCCTAGGGGGCGCTGCCAGGGCLCTGGCaCATGGLGTCCGGGTTCTGGAGGACGGCGTGAA
115
              T10
                      428 CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCCTGGAGGACGGCGTGAA
116
              SW2
122
              HK4
                      428 CCTTAGGGGGCGLTGCCAGAGCCCTGGCaCATGGLGTCCGGGTTgTGGAGGACGGCGTGAA
                      428 CETTAGGGGGGCGTGCCAGGGCCTTGGCGCATGGCGTCCGGGTTCTGGAAGACGCGTGAA
109
              SA10
110
               S45
                      428 CCCTAGGGGGCGCTGCCAGAGCCTTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
                      428 CCCTAGGGGGCGTTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTgTGGAGGACGGCGTGAA
123
               P8
                      428 CCLTAGGGGGCGTTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
124
               T3
120
               HK3
                      428 CCCTAGGGGGCGTTGCCAGAGCCLTGGCACATGGTGTCCGGGTTCTGGAGGACGGCGTGAA
                      428 CCCTAGGGGGCGTTGCCAGAGCCCTGGCACACGGTGTCCGGGTTCTGGAGGACGGCGTGAA
              HK5
121
108
              DR4
                      428 CCCTtGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGaGTTCTGGAAGACGGCGTGAA
104
             US11
                      428 CECTCGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
              S14
                      428 CcCTCGGqGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
105
                      428 CTCTLGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
106
               SW1
               S18
                      428 CTCTcGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
107
103
               DK7
                     428 CTCTEGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
103-124
       consensus
                          CccTaGGGGGGCGCTGCCAGGGCCCTGGCGCAtGGCGTCCGGGTtcTGGAGGACGGCGTGAA
SEO ID NO: ISOLATE
119
               59
                      117
              IND3
                      489 CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTAGCTTTGCTATCC
118
             TND8
                      489 CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
111
               D1
                      489 LTATGCAACAGGGAALTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCC
              US6
                      489 CTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCC
112
                      489 CTATGCAACAGGGAATCTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCC
              P10
113
114
              DK1
                      489 CTACGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTCTGLTGTCC
115
              TIO
                      489 CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTTTCTATCTTCCTCTTGGCTCTGCTGTCL
116
              SW2
                      489 CTATGCAACAGGGAATCTGCCCGGTTGCTCCTTTTCTATCTTCCTCTTGGCTLTGCTGTCC
122
              HK4
                      489 CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTCTGCTGTCC
                      489 CTATGCAACAGGGAATTTGCCCGGTTGCcCTTTCTCTATCTTCCTCTTGGCTLTGCTGTCC
109
             SALO
110
              S45
                      489 CTATGCAACAGGGAATCTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTcTGCTGTCC
123
               PΩ
                     489 CTATGCAACAGGGAATCTGCCTGGTTGCTCTTTCTCTATCTTCCTtTTGGCTTTGCTGTCt
124
               T3
                      489
                         LTACGCAACAGGGAATTTGCCTGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCC
120
              HK3
                     489 CTALGCAACAGGGAATTTACCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTGTCC
121
              HK5
                      489 CTACGCAACAGGGAATATACCCGGTTGCTCTTTCTCTATCTTCCTTTTTGGCTTTGCTGTCC
                      489 CTATGCAACAGGGAATCTTCCTGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTCTCT
108
              DR4
             US11
104
                      489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTCCTaGCCCTGCTTTCT
105
              S14
106
              SW1
                      107
              S18
                      103
              DK7
                     489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTtTGGCCCTGCTCTCT
103-124
       consensus
                          CTALGCAACAGGGAALCTGCCCGGTTGCLCLTTCTCTATCTTCCTCLTGGCLLTGCTGTCC
SEO ID NO: ISOLATE
119
               SS
                      550 TGTTTGACCATCCCAGCTTCCGCT
117
             IND3
                      550 TGTTTGACCATCCCAGCTTCCGCT
118
             INDS
                      550 TGTTTGACCGTCCCAGCTTCCGCT
111
               D1
                      550 TGTTTGACCATCCCAGCTTCCGCT
112
              US6
                      550 TGTTTGACCATtCCAGCTTCCGCT
113
                      550 TGCCTGACCATCCCAGCGTCCGCT
              P10
114
              DK1
                      550 TGTLTGACCATCCCAGCTTCCGCC
115
              T10
                      550 TGTCTGACCATCCCAGCTTCCGCT
116
              SW2
                      550 TGTCTGACCATCCCAGCTTCCGCT
                      550 TGTTTGACCATCCCAGCTTCCGCT
122
              HK4
109
             SA10
                      550 TGTTTaACCATCCCAGCTTCCGCT
110
              S45
                      550 TGCTTGACCATCCCAGCTTCCGCT
123
               P8
                     550 TGtcTGACCATCCCAGCTTCCGCT
124
               T3
                     550 TGCTTGACCATCCCAGCTTCCGCT
```

550 TGCTTGACCACCCCAGCTTCCGCT

550 TGtcTGACCACCCCAGtTTCCGCT

HK3

HK5

120

121

108 104 105	DR4 US11 S14	550 550	TGCLTGACCGTGCCCGCaTCgGCC TGCCTGACTGTGCCCGCTTCAGCC TGCCTGACTGTGCCCGCTTCAGCC TGCCTGACAGTGCCCGCGTCAGCC
106 107 103	SW1 S18 DK7	550	TGECTGACEGTGCCCGCGTCAGCE TGCCTGACCGTGCCCGCETCgGCC
103-124	consensus		TGttTgACcatcCCaGctTCcGCt

# FIGURE 6D

SEO ID NO: ISOLATE   128	1 ATGAGCACAAETCCTAAACCTCAAAGAAAAACCAAAAGAAACACEAACCGTCGCCCACAAG 1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACEAACCGTCGCCCACAAG 1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACEAACCGTCGCCCACAAG 1 ATGAGCACAAATCCAAAACCCCAAAGAAAAACCAEAAAGAAACACAAACCGTCGCCCACAAG ATGAGCACAAATCCEAAAACCECAAAGAAAAACCAEAAAGAAACAC-AACCGTCGCCCACA-G
128 T2	62 ACGTTAAGTTtCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGCTGCCGCGCAGGGG
125 T4	62 ACGTTAAGTT-CCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGGGG
126 US10	62 ACGTTAAGTTECCGGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGGGG
127 T9	62 ACGTTAAGTT-CCCGGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGGGG
125-128 consensus	ACGTTAAGTT-CCGGGCGGCCAGATCGTTGGCGGAGTATACTTGtTGCCGCGCAGGGG
SEO ID NO: ISOLATE	
128 T2	123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCtCGTGGAAGG
125 T4	123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGATCCCAGCCACGTGGGAGG
126 US10 127 T9	123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGGGGTCCCAGCCACGTGGGAGG 123 CCCLAGGTTGGGTGTGCGCACGACAAGGAAGACTTCGGAGGGGTCCCAGCCACGTGGGAGG
127 19	123 CCCEAGG11GGG1GCGCACGACAAGGAAGAC11CGGAGCGG1CCCAGCCACG1GGGAGG
125-128 consensus	CCCcAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGGAGG
SEO ID NO: ISOLATE	
128 T2	184 CGCCAGCCCATCCCLAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGATA-C
125 T4	184 CGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGATAEC
126 US10	184 CGCCAGCCCATCCCCAAAGATCGGCGCCCCCTGGCAAGTCCTGGGGAAAACCAGGATACC
127 T9	184 CGCCAGCCCATCCCCAAAGATCGGCGCtCCACTGGCAAGTCCTGGGGAAAACCAGGATACC
125-128 Consensus	CGCCAGCCCATCCC-AAAGATCGGCGCtCCACTGGCAAGTCCTGGGGAAAACCAGGATA-C
SEO ID NO: ISOLATE	
128 T2	245 CCTGGCCCCTGTATGGGAATGAGGGGCTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG
125 T4	245 CCTGGCCCTGTATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG
126 US10	245 CETGGCCCTATATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCCGAGG
127 T9	245 CcTGGCCtCTATATGGGAATGAGGGACTCGGCTGGGCgGGATGGCTCCTGTCCCCCGAGG
125-128 consensus	CeTGGCCeCT-TATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCCGAGG
SEO ID NO: ISOLATE	
128 T2 125 T4	306 TTC::CGTC:CTC::TGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAA:GTGGGTAAAGTC 306 TTC:CCGTC::TGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGTAAGGTC
125 14 126 US10	306 TTCCCGTCCCTCGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGTAAGGTC
127 T9	306 TTCCCGTCCCTCTTGGGGCCCCAgTGAcCCCCGGCATAGGTCGCGCAACGTGGGTAAGGTC
125-128 consensus	TTCcCGTCCCTCtTGGGGCCCCAaTGAcCCCCGGCATAGGTCGCGCAAcGTGGGTAAGGTC
SEO ID NO: ISOLATE	
128 T2	367 ATCGATACCCTAACGTGCgGCtTTGCCGACCTCATGGGGTACaTCCCCGTCGTAGGCGcCC
125 T4 126 US10	367 ATCGATACCCTAACGTGCAGCCTTGCCGACCTCATGGGGTACGTCCCCGTCGTAGGCGGCC
126 US10	367 ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGATACATCCCCGTCGTGGGGGCCC 367 ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGTACATCCCCGTCGTAGGCGCCC
125-128 consensus	ATCGATACCCTAACGTGCgGCtTTGCCGACCTCATGGGGTACaTCCCCGTCGTaGGCGccC
SEO ID NO: ISOLATE 128 T2	428 CGcTtGGTGGtGTCGCCAGAGCTCTtGCGCATGGCGTGAGAGTCCTGGAGGACGGAGTTAA

# FIGURE 6D

125 T4 126 US10 127 T9	428 CGLTGGGGGCCCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGACGGGTTAA 428 CGCTTGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGGGTCCTGGAGGACGGGGTTAA 428 CGCTTGGTGGCGTLGCCAGAGCTCTCGCGCACGGCGTGAGGGTCCTGGAGGACGGGGTTAA
125-128 consensus	CGcTtGGTGGcGTcGCCAGAGCTCTcGCGCAtGGCGTGAGAGTCCTGGAGGACGGGGTTAA
SEO ID NO: ISOLATE	
128 T2	489 TTATGCAACAGGLAACTTACCCGGTTGCTCCTTTTCTATcTTCTTGCTAGCCCTgCTGTCC
125 T4	489 TTATGCAACAGGGAACTTACCLGGTTGCTCCTTTTCTATLTTCTTGCTGGCCCTACTGTCC
126 US10	489 TTATGCAACAGGGAACTTACCcGGTTGCTCCTTTTCTATCTTCTTGCTGGCCtTACTGTCC
127 T9	489 TTATGCAACAGGGAACCTACCtGGTTGCTCtTTTTCTATCTTCTTGCTGGCCcTACTGTCC
125-128 consensus	TTATGCAACAGGGAACETACC-GGTTGCTCCTTTTCTATCTTCTTGCTGGCCCTACTGTCC
SEO ID NO: ISOLATE	
128 T2	550 TGCATCACLATTCCgGTtTCaGCT
125 T4	550 TGCATCACCATTCCAGTCTCcGCT
126 US10	550 TGCATCACCATTCCAGTCTCTGCT
127 T9	550 TGCATCACCACTCCgGcCTCTGCT
125-128 consensus	TGCATCACcAtTCC-GtcTCtGCT

# FIGURE 6E

SEC ID NO: ISOLATE   131 DK1   132 SW3   133 DK8   129 T8   130 US1   129-133 CONSENSUS	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAATACAAACCGCCGCCCACAGG 1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAATACAAACCGCCGCCCACAGG 1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACACCAAAAGAAACACAAACCGCCGCCCACAGG 1 ATGAGCACAAATCCTAAACCTAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCACAGG 1 ATGAGCACAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCCACAGG ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAAACCAAAACCGCCGCCCCACAGG
SEO ID NO:   ISOLATE	62 ACGITAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG 62 ACGITAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGAGAGTTTACTTGCTGCCGCGCAGGGG 62 ACGITAAGTTCCCGGGTGGCGCCAGATCGTTGGCGAGTTTACTTGCTGCCGGCAAGGG 62 ACGTCAAGTTCCCGGGTGGGGCCCAGATCGTTGGCGGGTTTACTTGCTGCCGGCAGGGG 62 ACGTCAAGTTCCCGGGTGGGGGCAGATCGTTGGCGGGTTTACTTGCTGCCGGCAGGGG ACGTCAAGTTCCCGGGTGGCGGCAGATCGTTGGCGGGTTTACTTGCTGCCGCCAGGGG
SEO ID NO: ISOLATE 131 DK11 132 SW3 133 DK8 129 T8 130 US1 129-133 Consensus	123 CCCCAGGTTGGGTGTGCGCACGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA 123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA 123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA 123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA 123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
SEO ID NO: ISOLATE   131 DK11   132 SW3   133 DK8   129 T8   130 US1   129-133 Consensus	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAG-CCTGGGGAAAGCCAGGATATC 184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAGCCAGGATATC 184 CGCCAGCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAACCAGGATATC 184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAACCAGGATATC 184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAAGCCAGGATATC CGCCAGCCCATCCCGAAAGATCGGGCGCTCCACCGGCAAGTCCTGGGGAAAAGCCAGGATATC
SEO ID NO: ISOLATE   131 DK11   132 SW3   133 DK8   129 T8   130 US1   129-133 Consensus	245 CTTGGCCCCTGTATGGAAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCCGCGG 245 CTTGGCCCCTGTATGGAAACGAGGGCTGGGGCTGGGCAGGTTGGCTCCTGTCCCCCCGCGG 245 CTTGGCCCCTGTATGGAAACGAGGGCTGGGGTTGGCAGGTTGGCTCTGTCCCCCCGCGG 245 CTTGGCCCTGTATGGAAACGAGGGCTTGGGCTGGGCAGGTTGGCTCTGTCCCCCCCGCG 246 CTTGGCCTCTGTACGGAAACGAGGGCTGGGCTTGGCGTTGGCTCTGTCCCCCCGCGG CTTGGCCCTGTALGGAAACGAGGGCTGCGGCTGGGCAGGTTGGCTCTGTCCCCCCCGCGG
SEC ID NO: ISOLATE 131 DK11 132 SW3 133 DK8 129 T8 130 US1 129-133 CONSENSUS	306 GTCTCATCCTAATTGGGGCCCCACTGACCCCCGGCATAAATCACGCAATTTGGGCAAAGTC 306 GTCTCATCCTAATTGGGGCCCCACTGACCCCCGGCATAGATCACGCAATTTGGGCAAAGTC 306 GTCTCGTCCTACTTGGGGCCCCACTGACCCCCGGGCATAGATCACGCAATTTGGGCAAAGTC 306 GTCTCGTCCTACTTGGGGCCCCACTGACCCCCGGGCATAGATCACGCATAATTTGGGCAAGGTC GTCTCGTCCTACTTGGGGCCCCACTGACCCCCGGCACAGATCACACAAATTTGGGCAAGTC GTCTCGTCCTACTTGGGGCCCCACTGACCCCCGGCACAGATCACCAAATTTGGGCAAGTC
SEC ID NO:         ISOLATE           131         DK11           132         SW3           133         DK8           129         T8	367 ATCGACACCATTACGTGTGGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTGGGCGCCC 367 ATCGACACCATTACGTGTGGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC 367 ATCGACACCATTACGTGTGGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC 367 ATCGATACCATTACATGTGGTTTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC

### FIGURE 6E

	130	US1	367	ATCGATACCATTACGTGTGTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC
	129-133	consensus		${\tt ATCGACACCATTACGTGTGGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTtGGCGCCC}$
	SEO ID NO			
	131	DK11		CGGTCGGAGGCGTCGCCAGAGCTCTTGGCACACGGTGTTAGAGTCCTGGAAGACGGGATAAA
	132	SW3		CGGTCGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTTAGAGTCCTGGAAGACGGGATAAA
	133	DK8		CGGTLGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTTAGGGTCCTGGAAGACGGGATAAA
	129	Т8		CGGTCGGAGGCGTCGCCAGAGCTCTGGCACALGGTGTTAGGGTCCTGGAAGACGGGATAAA
	130	US1	428	CGGTCGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTTAGGGTCCTGGAAGACGGGATAAA
	129-133	consensus		$\tt CGGTcGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTTAGGGTCCTGGAAGACGGGATAAA$
	SEO ID NO			
	131	DK11		TTACGCAACAGGGAATCTGCCTGGTTGCTCTTTTTCTATCTTCTTACTTGCTCTTCTGTCa
	132	SW3		TTACGCAACAGGGAATCTGCCTGGTTGCTCTTTTTCTATCTTCTTACTTGCTCTTCTGTCG
	133	DK8	489	TTACGCAACAGGGAATTTGCCTGGTTGCTCTTTTTTCTATCTTCTTGCTTG
	129	T8	489	CTALGCAACAGGGAATTTGCCTGGTTGCTCTTTTTCTATCTTCTTGCTTG
	130	US1	489	tTAcGCAACAGGGAATcTGCCTGGTTGCTCcTTTTCTATCTTCTTACTTGCTCTTCTGTCg
eş.	129-133	consensus		tTACGCAACAGGGAATCTGCCTGGTTGCTCtTTTTCTATCTTCTTaCTTGCTCTTCTGTCq
er Ke				
r.	SEO ID NO			
į	131	DK11	550	TGCTgCACAGTGCCAGTGTCTGCG
12	132	SW3		TGCTECACAGTGCCAGTGTCTGCG
174	133	DK8	550	TGCTgCACAGTGCCAGTGTCTGCG
5	129	T8		TGCTECACAGTGCCAGTGTCTGCA
4	130	US1	550	TGCgcCACgGTGCCgGTGTCTGCA
	129-133	consensus		TGCt-CACaGTGCCaGTGTCTGCq

SEO ID NO: ISOLATE	
131 DK11	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAATACAAACCGCCGCCCACAGG
132 SW3	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAATACAAACCGCCGCCCACAGG
133 DK8	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCACAGG
129 T8C	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCACAGG
130 US1	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCACAGG
125 T4	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAGG
126 US10	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACLAACCGTCGCCCACAAG
127 T9	1 ATGAGCACAAATCC&AAACCCCAAAGAAAAACCALAAGAAACACCAACCGTCGCCCACAGG
128 T2	1 ATGAGCACAALTCCTAAACCTCAAAGAAAAACCAAAAGAAACACTAACCGTCGCCCACAAG
	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACTAACCGcCGCCCACAGG
134 S83	1 ATGAGACAM TECTAMENTAL CONTROL OF THE CONTROL OF T
125-134 consensus	ATGAGCACAAATCCLAAACCLCAAAGAAAAACCAAAAGAAAACACAAACCGCCGCCCACAGG
SEO ID NO: ISOLATE	
131 DK11	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
132 SW3	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGGTTTACTTGCTGCCGCGCAGGGG
133 DK8	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
129 T8	62 ACGTCAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
130 US1	62 ACGTCAAGTTCCCGGGTGGCGGECAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
	62 ACGTTAAGTTCCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGGGG
	62 ACGTTAAGTTECCGGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGGGG
126 US10	
	62 ACGTTAAGTTcCCGGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGGGG
pars 128 T2	62 ACGTTAAGTTtCCGGGCGGCCAGATCGTTGGCGGAGTATACTTGCTGCCGCGCAGGGG
	62 ACGTCAAGTTcCCGGGCGGLGGCCAGATCGTTGGCGGAGTATACTTGCTGCCGCGCAGGGG
134 303	
125-134 consensus	ACGTLAAGTTcccGGG-GGcGGcCAGATCGTTGGCGGAGT-TACTTGcTGCCGCGCAGGGG
3	
SEO ID NO: ISOLATE	
131 DK11	123 CCCCAGGTTGGGTGTGCGCACGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
132 SW3	123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
	123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGtCTTCCGAGCGATCCCAGCCGCGTGGGAGG
1.77	123 CCCLAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
	123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
	123 CCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGATCCCAGCCACGTGGGAGG
	123 CCCCAGGITGGGGGGGACAAGGAAGCITCGGAGCGATCCCAGCCAGCGAGG
(f) 126 US10	123 CCCCAGGTTGGGTGTGCGCGCACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGGAGG
127 T9	123 CCCLAGGTTGGGTGTGCGCACGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGGAGG
128 T2	123 CCCcAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCLCGTGGaAGG
134 S83	123 CCCGAGATTGGGTGTGCGCGCGACGAGGAAAACTTCcGAACGGTCCCAGCCACGTGGGAGG
125-134 consensus	${\tt CCCcAGGTTGGGTGTGCGCgCGACaAGGAAgaCTTCcGAgCGaTCCCAGCCgCGTGGgAGg}$
SEO ID NO: ISOLATE	
131 DK11	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCCCTGGGGAAAGCCAGGATATC
132 SW3	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAGCCAGGATATC
132 SW3	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAACCGGGATATC
	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAACCAGGATATC
129 T8	
130 US1	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAGCCAGGATATC
125 T4	184 CGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGATATC
126 US10	184 CGCCAGCCCATCCCCAAAGATCGGCGCCCCCCTGGCAAGTCCTGGGGAAAACCAGGATACC
127 T9	184 CGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGATACC
128 T2	184 CGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGATACC
134 S83	184 CGCCAGCCCATCCCTAAAGATCGGCGCaCCACTGGCAAGTCCTGGGGAAggCCAGGATACC
125-134 consensus	CGCCAGCCCATCCCqAAAGATCGGCGCtCCAC-GGCAAGtCCTGGGGAAaaCCaGGATAtC
SEO ID NO: ISOLATE	
131 DK11	245 CTTGGCCCCTGTATGGAAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCCGCGG
132 SW3	245 CTTGGCCCCTGTATGGAAACGAGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCCGCGG
133 DK8	245 CTTGGCCCCTGTATGGAAACGAGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCCGCGG
129 T8	245 CTTGGCCTCTtTACGGAAACGAGGGCTGCGGtTGGGCAGGTTGGCTCCTGTCCCCCGCGG

84890\_1

# FIGURE 6F

	FIGURE OF
130 US1	245 CTTGGCCTCTGTACGGAAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCCGCGG
125 T4	245 CCTGGCCCCTGTATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCCGAGG
126 US10	245 CETGGCCCTATATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG
127 T9	245 CCTGGCCtCTATATGGGAATGAGGGACTCGGCTGGGCGGGATGGCTCCTGTCCCCCGAGG
128 T2	245 CCTGGCCCCTGTATGGGAATGAGGGGCTCGGCTGGGCAGGATGGCTCCTGTCCCCCCGAGG
134 S83	245 CtTGGCCCCTGTATGGGAATGAGGGCCTCGGCTGGGCAGGGTGGCTCCTGTCCCCCCGGGG
134 303	243 0010000010111100000110000100041000410
125-134 consensus	CtTGGCCcCTqTAtGG-AA-GAGGGcCGGcTGGGCaGGtTGGCTCCTGTCCCCCCGcGG
SEO ID NO: ISOLATE	
131 DK11	306 GTCTCATCCTAATTGGGGCCCCACTGACCCCCGGCATA&ATCACGCAATTTGGGLAAAGTC
132 SW3	306 GTCTCATCCTAATTGGGGCCCCACTGACCCCCGGCATAGATCACGCAATTTGGGCAAAGTC
133 DK8	306 GTCTCGTCCTACTTGGGGCCCCACTGACCCCCGGCATAGATCACGCAATTTGGGCAAAGTC
129 T8	306 GTCTCGTCCTACTTGGGGCCCCACTGACCCCCGGCATAGATCACGTAATTTGGGCAqAGTC
130 US1	306 GTCTCGTCCTACTTGGGGCCCCACTGACCCCCGGCACAGATCACGTAACTTGGGCAAGGTC
125 T4	306 TTCCCGTCCTCcTGGGGCCCCAaTGACCCCCGGCATAGGTCGCGCAACGTGGGTAAGGTC
126 US10	306 TTCCCGTCCTTTGGGGCCCCAcTGAtCCCCGGCATAGGTCGCGCAACGTGGGTAAGGTC
127 T9	306 TTCCCGTCCCTCTTGGGGCCCCAgTGACCCCCGGCATAGGTCGCGCAACGTGGGTAAGGTC
128 T2	306 TTCTCGTCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAALGTGGGTAAAGTC
134 S83	306 TTCTCGcCCtTCaTGGGGCCCCAccGACCCCGGCATAaaTCGCGCAActTGGGTAAgGTC
	•
125-134 consensus	-TCtCgtCCt-ctTGGGGCCCCActGAcCCCCGGCAtAgaTC-CGcAA-tTGGGtAa-GTC
SEO ID NO: ISOLATE	
131 DK11	367 ATCGACACCATTACGTGTGGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTcGGCGCCC
132 SW3	367 ATCGACACCATTACGTGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC
133 DK8	367 ATCGACACCATTACGTGTTTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC
129 T8	367 ATCGATACCATTACATGTGGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC
130 US1	367 ATCGATACCATTACGTGTGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC
125 T4	367 ATCGATACCCTAACGTGCAGCCTTGCCGACCTCATGGGGTACGTCCCCGTCGTAGGCGGCC
126 US10	367 ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGATACATCCCCGTCGTgGGCGCtC
127 T9	367 ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTAGGCGCCC
128 T2	367 ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTAGGCGCCC
134 S83	367 ATCGATACCCTAACGTGCGGETTTGCCGACCTCATGGGGTACATACCCGTCGTLGGCGCLC
125-134 consensus	ATCGAtACC-T-ACgTG-gGttTTGCCGACCTCATGGGgTACaTcCC-GTCGTtGGCGccC
SEO ID NO: ISOLATE	
131 DK11	428 CGGTCGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTTAGAGTCCTGGAAGACGGGATAAA
132 SW3	428 CGGTCGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTTAGAGTCCTGGAAGACGGGATAAA
133 DK8	428 CGGTtGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTTAGGGTCCTGGAAGACGGGATAAA
129 T8 130 US1	428 CGGTCGGAGGCGTCGCCAGAGCTCTCGCACALGGTGTTAGGGTCCTGGAAGACGGGATAAA 428 CGGTCGGAGGCGTCGCCAGAGCTCTCGCACACGGTGTTAGGGTCCTCGAAGACGGGATAAA
130 US1 125 T4	428 CGETGGGGGGTCGCCAGAGCTCTCGCACACGGTGTTAGGGTCCTGGAGACGGGGTTAAA 428 CGETGGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGACGGGGTTAA
125 14 126 US10	428 CGCTTGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGGGTCCTGGAGGACGGGGTTAA 428 CGCTTGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGGGTCCTGGAGGACGGGGTTAA
126 US10	428 CGCTTGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGGGTCCTGGAGGACGGGGTTAA
127 128 T2	428 CGCTTGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGACGGGGTTAA
128 134 S83	428 CGCTTGGTGGEGTEGCCAGAGCTCTEGCGCATGGCGTGAGAGTCCTGGAGGACGGGGTTAA 428 CGGTTGGCGGCGTEGCCAGAGCCCTCGCCCCATGGGGTGAGGGTECTGGAGGACGGGATAAA
151	120 ccg1100c00c31t0ccnoxoccc1cocccn100g01tc10dx0cdxcooga1axx
125-134 consensus	CggTtGGaGGcGTcGCCAGAGCtCTgGCaCA-GGtGT-AG-GTcCTGGA-GACGGgaTaAA
125 157 00000000	egg. to decomposition and the contract of the
SEO ID NO: ISOLATE	
131 DK11	489 TTACGCAACAGGGAATCTGCCTGGTTGCTCTTTTTCTATCTTCTTACTTGCTCTTCTGTCa
132 SW3	489 TTACGCAACAGGGAATCTGCCTGGTTGCTCTTTTTCTATCTTCTTACTTGCTCTTCTGTCG
133 DK8	489 TTACGCAACAGGGAATTTGCCTGGTTGCTCTTTTTCTATCTTCTTGCTTG
129 T8	489 CTALGCAACAGGGAATTTGCCTGGTTGCTCTTTTTCTATCTTCTTGCTTG
130 US1	489 TTACGCAACAGGGAATCTGCCTGGTTGCTCCTTTTCTATCTTCTTTACTTGCTCTTCTGTCG
125 T4	489 TTATGCAACAGGGAACTTACCTGGTTGCTCCTTTTCTATETTCTTGCTGGCCCTACTGTCC
126 US10	489 TTATGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCCLTACTGTCC
127 T9	489 TTATGCAACAGGGAACCTACCLGGTTGCTCLTTTTTCTATCTTCTTGCTGGCCCTACTGTCC
128 T2	489 TTATGCAACAGGLAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTAGCCCTGCTGCCC
	3

#### FIGURE 6F

134	S83	489	${\tt TTATGCAACgGGgAAtTTgCCCGGTTGCTCtTTcTCTATCTTtcTcttgGCCCTctTgTCt}$
125-134	consensus		$\verb ttatgCAACaGGgAAtttgCCtGGTTGCTCtTttTCTATcTTcttgcttgC-cTtcTGTCc \\$
SEO ID NO	o: ISOLATE		
131	DK11	550	TGCTgCACAGTGCCAGTGTCTGCG
132	SW3		TGCTECACAGTGCCAGTGTCTGCG
133	DK8	550	TGCTqCACAGTGCCAGTGTCTGCG
129	T8	550	TGCTECACAGTGCCAGTGTCTGCA
130	US1	550	TGCgcCACgGTGCCgGTGTCTGCA
125	T4	550	TGCATCACCATTCCAGTCTCcGCT
126	US10	550	TGCATCACCATTCCAGTCTCTGCT
127	<b>T9</b>	550	TGCATCACCACTCCGGCCTCTGCT
128	T2	550	TGCATCACTATTCCGGTTTCAGCT
134	S83	550	TGCATCtCTgTgCCaGTTTCcGCc
125-134	consensus		TGCatCaCagtgCCaGtgTCtGCt

# FIGURE 6G

SEO ID NO:         ISOLATE           138         DK12           135         HK10           136         S52           137         S2	1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAAGAAACACCATCCGTCGCCCACAGG 1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAAGAAACACCATCCGTCGCCCACAGG 1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAAGAAAACACCATCCGTCGCCCACAGG 1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAAGAAAACACCATCCGTCGCCCCACAGG
135-138 consensus	ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAAGAAACACCATCCGTCGCCCACAGG
SEO ID NO:         ISOLATE           138         DK12           135         HK10           136         S52           137         S2           135-138         consensus	62 ACGTCAAGTTCCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG 62 ACGTTAAGTTCCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG 62 ACGTTAAGTTCCCGGGTGGCGGACGAGTGGTTGGGTGGAGTATACGTGTTTGCCGCCAGGGG 62 ACATCAAGTTCCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG ACGT-AAGTTCCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG
SEO ID NO:         ISOLATE           138         DK12           135         HK10           136         S52	123 CCCACGATTGGGTGTGCGCGCGACGCGTAAAACTTCTGAACGGTCaCAGCCTCGCGGACGg 123 CCCACGATTGGGTGTGCGCGCACGCGTAAAACTTCTGAACGGTCGAGAGCTGCGGGACGA 123 CCCACGATTGGGTGTGCGGCGCAGACGCTAAAACTTCTGAACGGTCACAGCCTGCGGAGCA
137 S2	123 CCCACGATTGGGTGTGCGCGCGCGCGTAAAACTTCTGAACGGTCACAGCCTCGCGGACGG CCCACGATTGGGTGTGCGCGCGACGCGTAAAACTTCTGAACGGTCACAGCCTCGCGGACG-
\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$\\$	184 CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGGTCCTGGGCTCAGCCLGGGTACC 184 CGACAGCCTATCCCCCAAGGCGCGTCGGAGCGAAGGCCGGTCCTGGGTTCAGCCCGGGTACC 184 CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGGTCCTGGGCTCAGCCCGGGTACC 184 CGACAGCCTATCCCCAAGGCGGTCGGAGCGAAGGCCGATCCTGGGCTCAGCCCGGGTACC
135-138 Consensus	CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGgTCCTGGGCTCAGCCcGGGTACC
SEO ID NO.   ISOLATE	245 CTTGGCCCCTCTATGGTAACGAGGGCTGCGGGTGGGCAGGGTGGCTCCTGTCCCCACGCGG 245 CTTGGCCCCTCTATGGTAACGAGGGCTGCGGGTGGCAGGTGGCTCCTGTCCCCACGCGG 245 CTTGGCCCCTCTATGGTAACGAGGCTGCGGGTGGCAGGTGGCTCCTGTCCCCACGCGG 245 CTTGGCCCCTCTATGGTAACGAGGCTGCGGTGGGCAGGGTGGCTCCTGTCCCACGCGG CTTGGCCCCTCTATGGTAACGAGGGCTGCGGTGGGCAGGGTGGCTCCTGTCCCCACGCGG
SEO ID NO:         ISOLATE           138         DK12           135         HK10           136         S52           137         S2           135-138         consensus	306 CTCCCGTCCATCTTGGGGCCCAAACGACCCCCGGCGGAGGTCCCGCAATTTGGGTAAGGTC 306 CTCCCGTCCATCTTGGGGCCCAAACGACCCCCGGCGGAGGTCCCGCAATTTGGGTAAAGTC 306 CTCCCGTCCATCTTGGGGCCCAAACGACCCCCGGGGGGGG
SEQ ID NO:         ISOLATE           138         DK12           135         HK10           136         S52           137         S2           135-138         consensus	367 ATCGATACCCT-ACGTGCGGATTCGCCGACCTCATGGGTACATCCCGCTCGTCGGCGCTC 367 ATCGATACCCTTACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC 367 ATCGATACCCTTACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC 367 ATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC ATCGATACCCTTACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC
SEQ ID NO: ISOLATE 138 DK12	428 CLGTAGGgGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA

#### FIGURE 6G

135 HK1	0 428	CCGTAGGAGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
136 S5	2 428	CCGTAGGAGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
137 S	2 428	CCGTAGGAGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
135-138 consensu	6	CcGTAGGAGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
SEO ID NO: ISOLAT		
138 DK1	2 489	TTTCGCAACAGGGAACTTGCCCGGTTGCTCCTTTTCTATCTTCCTTC
135 HK1	0 489	TTTCGCAACAGGGAACTTGCCCGGTTGCTCCTTTTCTATCTTCCTTC
136 S5	2 489	TTTTGCAACAGGGAACTTGCCCGGTTGCTCCTTTTCTATCTTCCTTC
137 S	2 489	TTTTGCAACAGGGAACTTGCCCGGTTGCTC+TTTTCTATCTTCCTTCTTGCcCTGTTCTC+
135-138 consensu	s	TTT-GCAACAGGGAACTTGCCCGGTTGCTCcTTTTCTATCTTCCTTCTTGCtCTGTTCTCt
SEO ID NO: ISOLAT		
138 DK1	2 550	TGCCTAATTCATCCAGCAGCTAGT
135 HK1	0 550	TGCTTAATTCATCCAGCAGCTAGT
136 S5	2 550	TGCTTAgTTCATCCtGCAGCTAGT
137 S	2 550	TGCTTAaTTCATCCAGCAGCTAGT
135-138 consensu	B	TGCtTAaTTCATCCaGCAGCTAGT

# FIGURE 6H

SEO ID NO:	ISOLATE	
145		1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCAATGG
143	Z6	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCCAAACGTAACACCCAACCGCCGCCCCATGG 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCCAACCGCCCCCATGG
144		1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCGAACCGCCCCCATGG
140		1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCCATGG
139		1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCGACCG
142		1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCCATGG
141	Z1	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCCCCATGG
		ATOMS CHEATHER CONTROL AND CONTROL AND CONTROL
139-145 co	onsensus	ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGcCGCCCCATGG
SEO ID NO:	ISOLATE	
145	DK13	52 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGTTGCCGCGCAGGGG
143	26	52 ACGTTAAGTTCCCGGGTGGTGGCCAGATCGTTGGCGGAGGTTTACTTGTTGCCGCGCAGGGG
144	27 (	52 ACGTTAAGTTCCCGGGCGGTGGCCAGATCGTTGGCGGAGGTTTACTTGTTGCCGCGCAGGGG
140	Z8 (	52 AtGTAAAaTTCCCaGGCGGCGGCCAGATCGTTGGCGGAGGTTTACTTGTTGCCGCGCAGGGG
139		2 AcGTAAAgTTCCCGGGTGGTGGCCAGATCGTTGGCGGAGTTTACTTGTTGCCGCGCAGGGG
142	25	52 ATGTAAAATTCCCGGGTGGTGGLCAGATCGTTGGCGGAGGTTTACTTGTTGCCGCGCAGGGG
141	Z1 (	52 ATGTGAAATTCCCGGGCGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
139-145 cc	nsensus	Acgt-Aagttcccgggtggtggccagatcgttggcggagtttacttgttgccgcgcagggg
		THE THE PERSON OF THE PERSON O
SEO ID NO:	ISOLATE	
145		23. 000-10. mmccommonosco
143		23 CCCLAGaTTGGGTGTGCGCGCGACTAGGAAGACTTCGGAGCGGTCGCAACCTCGTGGGAGG
144		23 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCGGAGCGGTCGCAACCTCGTGGGAGA
140		23 CCCCAGaTTGGGTGTGCGCACCACTAGGAAGACTTCGGAGCGGTCGCAACCTCGTGGGAGA
139		23 CCCCAGGTTGGGTGTGCGCGCACTCGGAAGACTTCGGAGCGGTCGCAACCTCGTGGCAGG
142		23 CCCCAGGTTGGGTGTGCGCGCACTCGAAGACTTCGGAGCGGTCGCAACCTCGTGGCAGG
141		23 CCCCAGGTTGGGTGTGCGCGCACTCGGAAGACTTCGGAGCGGTCGCAACCTCGCGGCAGG
	21 1	23 CCCCcGGTTGGGTGTGCGCGCagCTCGGAAGACTTCGGAGCGGTCaCAACCTCGtGGCAGG
139-145 co	nsensus	CCCcaGgTTGGGTGTGCGCgCgaCTcGgAAGACTTCGGAGCGGTCgCAACCTCGtGGcAGg
139-145 co	nsensus	${\tt CCCcaGgTTGGGTGTGCGCgCgaCTcGgaAGACTTCGGAGCGGTCgCaACCTCGtGGcAGg}$
		CCCcaGgTTGGGTGTGCGCgCgaCTcGgAAGACTTCGGAGCGGTCgCAACCTCGtGGCAGg
	ISOLATE	•
SEQ ID NO:	ISOLATE DK13 1	84 CGCCAGCCTATCCCCAAGGCgCGcCaActeGAGGGLAGGTCCTGGGCTCAGCCtGGGTATC
SEO ID NO:	ISOLATE DK13 1 Z6 1	84 CGCCAGCCTATCCCCAAGGCgCGCCAACtcGAGGCLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGAAGGTCCTGGGCTCAGCCCGGGTATC
SEO ID NO: 145 143	ISOLATE DK13 1 Z6 1 Z7 1	84 CGCCAGCCTATCCCCAAGGCgCGcCaActeGAGGGLAGGTCCTGGGCTCAGCCtGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGAAGGTCCTGGGCTCAACCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTAGCCGGGTCACCCGGGTACC
SEQ ID NO: 145 143 144	ISOLATE DK13 1 Z6 1 Z7 1 Z8 1	CGCCAGCCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC     CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGAAGGTCCTGGGCTCAGCCCGGGTATC     CGTCAGCCTATCCCCAAGGCAGGTCGTATCTGAGGAAGGTCCTGGGCTCAACCCGGGTACC     CGCCAGCCTATCCCCAAGGCAGGTGCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC     CGCGGCCTATCCCCAAGGCAGGTGGTCGAGGGAAGGTCAGCCCGGGTACC
SEO ID NO: 145 143 144 140	ISOLATE DK13 1 Z6 1 Z7 1 Z8 1 Z4 1	84 CGCCAGCCTATCCCCAAGGCgCGcCaActcGAGGGLAGGTCCTGGGCTCAGCCtGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 84 CGTCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCAACCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCGAGGGAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCCGAGGGAGGTCCTGGGCCAGCCCTGGGTACC
SEO ID NO: 145 143 144 140 139	ISOLATE  DK13 1 26 1 Z7 1 Z8 1 Z4 1 Z5 1	CGCCAGCCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC     CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGAAGGTCCTGGGCTCAGCCCGGGTATC     CGTCAGCCTATCCCCAAGGCAGGTCGTATCTGAGGAAGGTCCTGGGCTCAACCCGGGTACC     CGCCAGCCTATCCCCAAGGCAGGTGCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC     CGCGGCCTATCCCCAAGGCAGGTGGTCGAGGGAAGGTCAGCCCGGGTACC
SEO ID NO: 145 143 144 140 139 142 141	ISOLATE  DK13 1  Z6 1  Z7 1  Z8 1  Z4 1  Z5 1  Z1 1	84 CGCCAGCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCTCGATCTGAGGGAAGGTCCTGGGCTCAGCCCGGGTATC 84 CGTCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGTCGGAGGGAAGGTCCTGGGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCGCGCGAGCAGAGGGCAAGTCCTGGGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCGCGCGGGTCGAGGCGAGGTCCTGGCCGGGTACC 84 CGTCAGCCTATCCCCCAAGGCGCCGGGTCGAGGCGAGGTCCTGGCCGGGTACC 84 CGTCAGCCTATCCCCCAAGGCGCCGGGTCCGAGGGCCGGGTCCGACCGGGTACC
SEO ID NO: 145 143 144 140 139 142 141	ISOLATE  DK13 1 26 1 Z7 1 Z8 1 Z4 1 Z5 1	84 CGCCAGCTATCCCCAAGGCGCGAACLGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCACCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCACCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCAGGTCGAGGGAGGAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCAGGGGCAGCAGAGGGCAGATCCTGGGCGAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCAGCGGGTCCGAGGGAGGTCCTGGGCGGGTCACCCGGGTACC
SEO ID NO: 145 143 144 140 139 142 141 139-145 co	ISOLATE  DK13 1 26 1 27 1 28 1 24 1 25 1 25 1 21 1 nsensus	84 CGCCAGCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCTCGATCTGAGGGAAGGTCCTGGGCTCAGCCCGGGTATC 84 CGTCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGTCGGAGGGAAGGTCCTGGGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCGCGCGAGCAGAGGGCAAGTCCTGGGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCGCGCGGGTCGAGGCGAGGTCCTGGCCGGGTACC 84 CGTCAGCCTATCCCCCAAGGCGCCGGGTCGAGGCGAGGTCCTGGCCGGGTACC 84 CGTCAGCCTATCCCCCAAGGCGCCGGGTCCGAGGGCCGGGTCCGACCGGGTACC
SEO ID NO: 145 143 144 140 139 142 141 139-145 CO	ISOLATE  DK13 1 26 1 27 1 28 1 24 1 25 1 21 1  nsensus  ISOLATE	84 CGCCAGCCTATCCCCAAGGCGCGCCAACTCGAGGGTAGGCTCTGGGCTCAGCCTGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCTCGATCTGAGGGAAGGTCCTGGGCTCAGCCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGGAAGGTCCTGGGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCCGAGGGGAGGGCTCAGCCCGGGTACC 84 CGTCAACCTATCCCCAAGGCACGCTCGGGCAGGCAGGTCCTGGGCGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGCTCGGGGCAGGTCCTGGGCGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCCAAGGCGCGGCCGGGCCCGGGTCCGACCGGGTACC CGCCAGCCTATCCCCCAAGGCGCCGGTCCGAGGGCAGGTCCTGGGCCTCAGCCCGGGTACC CGCCAGCCTATCCCCCAAGGCGCACGCCGGGCAGGCCCGGGTACC
SEO ID NO: 145 143 144 140 139 142 141 139-145 CO	ISOLATE  DK13 1 26 1 27 1 28 1 24 1 25 1 25 1 21 1  nsensus  ISOLATE DK13 2	84 CGCCAGCCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCGCTCGATCTGAGGGAAGGTCCTGGGCTCAGCCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGGCCAGAGCCGGGTACCCGGGTACCCAGCCCGGGTACCCAGCCCCGGGTACCCAGCCCCGGGTACCCAGCCCCGGGTACCCAGCCCCGGGTACCCAGCCCCGGGTACCCAGCCCCGGGTACCCAGCCCCGGGTACCCAGCCCTAGCCCAGCCCAGCCCCGGGTACCCCGGGTACCCCGGGTACCCAGCCCTATCCCCCAAGGCACGCCGGGTCCGAGGGCAGGTCCTGGGCCTAGCCCGGGTACCCCGGGTACCCAGCCCTATCCCCCAAGGCACGCCGGGTACCCAGCCCGGGTACCCAGCCCTATCCCCCAAGGCACGCCGGGTACCCCGGGTACCCCGGGTACCCAGCCCGGTACCCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCCGGGTACCCCCGGGTACCCCCGGGTACCCCGGGTACCCCGGGTACCCCCGGGTACCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCGGGTACCCCGGGTACGCACGACGACCACACCCCGGGTACCACACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGTACCCCGGGT
SEO ID NO: 145 143 144 140 139 142 141 139-145 co SEO ID NO: 145	ISOLATE  DK13 1 26 1 27 1 28 1 25 1 25 1 1 nsensus  ISOLATE  DK13 2 26 2	84 CGCCAGCCTATCCCCAAGGCGCGCCAACLEGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCGCACCCGATCTGAGGGAAGGTCCTGGGCTCAGCCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGGAAGGTCCTGGGCTCAACCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCGAGGGAGGCCTGGGTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCGCCGAGGCAGGAGGCAGTCCTGGGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCCAAGGCGCGGCGGCAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCCAAGGCGCGCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCCAAGGCGCAGCCGGGTCGGGGAGGCCTGGGCCCAGCCCGGGTACC  45 CLTGGCCCTTTACGGAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGGGGTGGGCGGATTGCTCCTGTCACCCCCTTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGGGGTGGGCGGATTGCTCCTGTCACCCCCTTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGGGGTGGGGGGATTGCTCCTGTCACCCCCTTGG
SEO ID NO: 145 143 144 140 139 142 141 139-145 CO SEO ID NO: 145 143	ISOLATE  DK13 1 26 27 1 28 1 24 1 25 1 21 1  nsensus  ISOLATE  DK13 2 26 2 27 2	84 CGCCAGCCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCGCCGCAACLCGAGGGAAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCAACCCGGGTATC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGCGGCAGCAGGGAGAGCCTGGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGCGGTCCGAGGGGAGGG
SEO ID NO: 145 143 144 140 139 142 141 139-145 co  SEO ID NO: 145 143 144 144	ISOLATE  DK13 1 26 1 27 28 1 28 24 1 25 21 1  INSERTING THE PROPERTY OF THE PR	84 CGCCAGCCTATCCCCAAGGCGCGCCAACLEGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCAGTCGTATCTGAGGAAGGTCCTGGGCTCAGCCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGAAGGTCCTGGGCTCAACCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCCGAGGGAAGGTCCTGGGCTCAGCCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCGCGCAGGCCAGGAGGCCTGGGCCAGGCCCAGGTACC 85 CGTCAGCCTATCCCCCAAGGCACGCCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCGCGCGCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCGCGCGCGGCGGAGGGCCTGGGCCCGGTCCC 45 CATGGCCTTTACGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAATGAGGGTTGCGGTGGGCAGGATGGCTCTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAATGAGGGTTGCGGTGGGCAGGATGGCTCTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAATGAGGGTTGCGGTGGGCAGGATGGCTCTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAACGAGGTTGCGGTGGGCAGGATGGCTCTGTTCACCCCGTGG 45 CATGGCCTCTTTACGGTAACGAGGGTTGCGGTTGGGCAGGATGGCTCTGTTCACCCCGTGG 45 CATGGCCTCTTTACGGTAACGAGGGTTGCGGTTGGGGCAGGATGGCTCTGTTCACCCCGTGG 45 CATGGCCTCTTTACGGTAACGAGGGTTGCGGTGGGCAGGATGGCTCTTGTACCCCCGTGG 45 CATGGCCTCTTTACGGTAACGAGGTTGCGGTGGGCAGGATGGCTCTTGTACCCCCGTGG 45 CATGGCCTCTTTTACGGTAACGAGGTTGCGGTTGGGGCAGGATGGCTCTTGTACCCCCGTGG 45 CATGGCCTCTTTACGGTAACGAGGTTGCGGTTGGGGCAGGATGGCTCTGTGACCCCGTGG 45 CATGGCCTCTTTACGGTAACGAGGTTGCGGTTGGGGCAGGATGGCTCTGTGACCCCGTGG 45 CATGGCCTCTTTACGGTAACGAGGTTGCGGTTGGGGCAGGATGGCTCTGTGACCCCCGTGG
SEO ID NO: 145 143 144 140 139 142 141 139-145 CO SEO ID NO: 145 143 144 140 139	ISOLATE  DK13 1 27 26 1 28 1 25 1 25 1 27 1 28 1 26 27 2 27 2 28 2 24 2 24 2	84 CGCCAGCCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCGCCGCAACLCGAGGGLAGGTCCTGGGCTCAGCCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCGAGGGAGGTCGCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGGCCAGAGGCGGGTACCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGCGGTCCGAGGGGAGGTCCTGGGCTCAGCCCGGGTACC 85 CGTCAGCCTATCCCCAAGGCACGTCCGAGGGGAGGTCCTGGGCTCAGCCCGGGTACC  CGCCAGCCTATCCCCAAGGCACGCGGTCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC  45 CLTGGCCCTTTACGGAATGAGGGCTGCGGGTGGGCGGAATGGCTCCTGTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGCGGGTGGGCGGAATGGCTCCTGTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGCGGGTGGGCAGAATGGCTCTGTTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGCGGGTGGGCAGAATGGCTCTGTTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAAGGGTTGCGGGTGGGCAGAATGGCTCTTGTCACCCCGTGG 45 CATGGCCTTTTACGGTAAAGAGGGTTGCGGGTGGGCAGAATGGCTCTTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAAAGAGGGTTGCGGGTGGGCAGAATGGCTCTTGTCACCCCGTGG 45 CATGGCCCTTTTACGGTAAAGAGGGTTGCGGGTGGGCAGAATGGCTCTTGTCCCCCGGTGG 45 CATGGCCCTTTTACGGTAAAGAGGGTTACGGGTGGGCAGAATGGCTCTTGTCCCCCGGTGG 45 CATGGCCCTCTTTACGGTAAAGAGGGTTACGGGTGGGCAGAATGGCTCTTGTTCCCCCCGTGG 45 CATGGCCCTCTTTACGGTAAAGAGGGTTACGGGCTGGGAGAATGGCTCTTGTTCCCCCCGTGG 45 CATGGCCCCTCTTACGGAATGAGGCTCCTGTTCCCCCCGTGG 45 CATGCCCCTCTTTACGGTAATGAAGGCTTCCGGGTGGGCAGAATGGCTCTTTTTCCCCCCGTGG 45 CATGCCCCTCTTTACGGTAATGAAGGCTTCCGGGCTGGGAATGGCTCCTTTTTCCCCCCGTGG 45 CATGCCCCTCTTTACGGTAATGAAGGCTTCGGGCTGGGAATGGCTCCTTTTCCCCCCGTGG 45 CATGCCCCCTCTTACGGTAATGAAGGCTTCGGGCTGGGAAGGCAGATGCTCTTTTCCCCCCGTGG 45 CTTTGCCCCCCTCTTTCCCCCCTTCCTTCCCCCCTGCTCTTTCCCCCC
SEO ID NO: 145 143 144 140 139 142 141 139-145 co  SEO ID NO: 145 144 144 140 139 144 144 140 139	ISOLATE  DK13 1 26 1 27 27 28 1 28 24 1 25 1 10 1  DR13 2 1  DR13 2 2 2 2 2 2 2 4 2 2 2 5 5	84 CGCCAGCCTATCCCCAAGGCgCGcCaActeGAGGGLAGGTCCTGGGCTCAGCCGGGTATC 85 CGCCAGCCTATCCCCAAGGCACCGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 86 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGGAAGGTCCTGGGCTCAACCGGGTATC 86 CGTCAGCCTATCCCCAAGGCACGTCGATCTGAGGAGGTCCTGGGCTCAACCCGGGTACC 86 CGTCAGCCTATCCCCAAGGCACGCGGCAGGCAGAGGCTCGGGCTACCCGGGTACC 87 CGTCAGCCTATCCCCAAGGCACCGCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 88 CGTCAGCCTATCCCCAAGGCGCGCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACGCCGGTCCGAGGGCAGGTCCTGGGCCCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACGCCGGCGGGCAGGGGCAGGTCCTCGTCACCCCGTGG 45 CATGGCCTTTTACGGAATGAGGGTTGCGGGTGGCAGGATGGCTCCTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAATGAGGGTTGCGGGTGGCAGGATGGCTCCTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAATGAGGGTTGCGGGTGGCAGGATGGCTCCTGTCACCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTGAGGGTGGCAGGATGGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTTAGGCAATGAGGGCTGAGGGGGAAGGGTGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTGAGGGGGAAGGGTGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTATGCCAATGAGGGCTGAGGGGGCAAGGGTGGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTGAGGGGGCAAGGGTGGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTGTGGGGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTTGGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTTGGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGGCCCCTTTATGCCAATGAGGGCTTGGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTTGGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGCCCCCTTTATGGCAATGAGGGCTTGTGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGCCCCCTTTATGCCAATGAGGGCTTGTGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGCCCCCTTTATGCCCAATGAGGCTTGTGGTGGCAAGGTGGCTCCTGTCCCCCCCGGG
SEO ID NO: 145 143 144 140 139 142 141 139-145 CO SEO ID NO: 145 143 144 140 139	ISOLATE  DK13 1 26 1 27 27 28 1 28 24 1 25 1 10 1  DR13 2 1  DR13 2 2 2 2 2 2 2 4 2 2 2 5 5	84 CGCCAGCCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCGCCGCAACLCGAGGGLAGGTCCTGGGCTCAGCCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGTCGAGGGAGGTCGCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGTCGGGCCAGAGGCGGGTACCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACGCGGTCCGAGGGGAGGTCCTGGGCTCAGCCCGGGTACC 85 CGTCAGCCTATCCCCAAGGCACGTCCGAGGGGAGGTCCTGGGCTCAGCCCGGGTACC  CGCCAGCCTATCCCCAAGGCACGCGGTCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC  45 CLTGGCCCTTTACGGAATGAGGGCTGCGGGTGGGCGGAATGGCTCCTGTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGCGGGTGGGCGGAATGGCTCCTGTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGCGGGTGGGCAGAATGGCTCTGTTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAGGGTTGCGGGTGGGCAGAATGGCTCTGTTCACCCCGTGG 45 CATGGCCTTTTACGGTAATGAAGGGTTGCGGGTGGGCAGAATGGCTCTTGTCACCCCGTGG 45 CATGGCCTTTTACGGTAAAGAGGGTTGCGGGTGGGCAGAATGGCTCTTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAAAGAGGGTTGCGGGTGGGCAGAATGGCTCTTGTCACCCCGTGG 45 CATGGCCCTTTTACGGTAAAGAGGGTTGCGGGTGGGCAGAATGGCTCTTGTCCCCCGGTGG 45 CATGGCCCTTTTACGGTAAAGAGGGTTACGGGTGGGCAGAATGGCTCTTGTCCCCCGGTGG 45 CATGGCCCTCTTTACGGTAAAGAGGGTTACGGGTGGGCAGAATGGCTCTTGTTCCCCCCGTGG 45 CATGGCCCTCTTTACGGTAAAGAGGGTTACGGGCTGGGAGAATGGCTCTTGTTCCCCCCGTGG 45 CATGGCCCCTCTTACGGAATGAGGCTCCTGTTCCCCCCGTGG 45 CATGCCCCTCTTTACGGTAATGAAGGCTTCCGGGTGGGCAGAATGGCTCTTTTTCCCCCCGTGG 45 CATGCCCCTCTTTACGGTAATGAAGGCTTCCGGGCTGGGAATGGCTCCTTTTTCCCCCCGTGG 45 CATGCCCCTCTTTACGGTAATGAAGGCTTCGGGCTGGGAATGGCTCCTTTTCCCCCCGTGG 45 CATGCCCCCTCTTACGGTAATGAAGGCTTCGGGCTGGGAAGGCAGATGCTCTTTTCCCCCCGTGG 45 CTTTGCCCCCCTCTTTCCCCCCTTCCTTCCCCCCTGCTCTTTCCCCCC
SEO ID NO: 145 143 144 140 139 142 141 139-145 CO SEO ID NO: 145 144 144 140 139 144 144 141	ISOLATE  DK13 1 26 1 27 27 28 1 28 24 1 25 1 10 1  DR13 2 1  DR13 2 2 2 2 2 2 2 4 2 2 2 5 5	84 CGCCAGCCTATCCCCAAGGCgCGcCaActeGAGGGLAGGTCCTGGGCTCAGCCtGGGTATC 85 CGCCAGCCTATCCCCAAGGCACCTCGATCTGAGGGAAGGTCCTGGGCTCAGCCAGGTATC 86 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGAAGGTCCTGGGCTCAACCCGGGTATC 87 CGTCAGCCTATCCCCAAGGCACGTCGATCTGAGGAAGGTCCTGGGCTCAACCCGGGTACC 88 CGTCAGCCTATCCCCAAGGCACGCAGGCACGAGGAGGTCCTGGGCTCAGCCCGGGTACC 80 CGTCAGCCTATCCCCAAGGCACCACCGGCCAGGAGGAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACGCCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACGCCGCGGCAGGGCAG
SEO ID NO: 145 143 144 140 139 142 141 139-145 CO SEO ID NO: 145 144 144 140 139 144 144 140 139 144	ISOLATE  DK13 1 26 1 27 27 28 1 28 24 1 25 1 1  INSERSUATE  DK13 2 26 2 27 2 28 2 24 2 25 2 21 2	84 CGCCAGCCTATCCCCAAGGCgCGcCaActeGAGGGLAGGTCCTGGGCTCAGCCGGGTATC 85 CGCCAGCCTATCCCCAAGGCACCGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 86 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGGAAGGTCCTGGGCTCAACCGGGTATC 86 CGTCAGCCTATCCCCAAGGCACGTCGATCTGAGGAGGTCCTGGGCTCAACCCGGGTACC 86 CGTCAGCCTATCCCCAAGGCACGCGGCAGGCAGAGGCTCGGGCTACCCGGGTACC 87 CGTCAGCCTATCCCCAAGGCACCGCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 88 CGTCAGCCTATCCCCAAGGCGCGCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACGCCGGTCCGAGGGCAGGTCCTGGGCCCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACGCCGGCGGGCAGGGGCAGGTCCTCGTCACCCCGTGG 45 CATGGCCTTTTACGGAATGAGGGTTGCGGGTGGCAGGATGGCTCCTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAATGAGGGTTGCGGGTGGCAGGATGGCTCCTGTCACCCCGTGG 45 CATGGCCTCTTTACGGTAATGAGGGTTGCGGGTGGCAGGATGGCTCCTGTCACCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTGAGGGTGGCAGGATGGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTTAGGCAATGAGGGCTGAGGGGGAAGGGTGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTGAGGGGGAAGGGTGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTATGCCAATGAGGGCTGAGGGGGCAAGGGTGGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTGAGGGGGCAAGGGTGGCTCCTGTCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTGTGGGGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTTGGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTTGGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGGCCCCTTTATGCCAATGAGGGCTTGGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGGCCCCTTTATGGCAATGAGGGCTTGGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGCCCCCTTTATGGCAATGAGGGCTTGTGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGCCCCCTTTATGCCAATGAGGGCTTGTGGGTGGCAAGGGTGGCTCCTGTCCCCCCGGGG 45 CTTGCCCCCTTTATGCCCAATGAGGCTTGTGGTGGCAAGGTGGCTCCTGTCCCCCCCGGG
SEO ID NO: 145 143 144 140 139 142 141 139-145 co SEO ID NO: 145 143 144 140 139 142 141 139-145 co	ISOLATE  DK13 1 26 1 27 2 28 1 25 1 27 1 25 1 27 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 29 22 24 2 20 25 2	84 CGCCAGCCTATCCCCAAGGCgCGcCaActeGAGGGLAGGTCCTGGGCTCAGCCtGGGTATC 85 CGCCAGCCTATCCCCAAGGCACCTCGATCTGAGGGAAGGTCCTGGGCTCAGCCAGGTATC 86 CGCCAGCCTATCCCCAAGGCACTCGATCTGAGGAAGGTCCTGGGCTCAACCCGGGTATC 87 CGTCAGCCTATCCCCAAGGCACGTCGATCTGAGGAAGGTCCTGGGCTCAACCCGGGTACC 88 CGTCAGCCTATCCCCAAGGCACGCAGGCACGAGGAGGTCCTGGGCTCAGCCCGGGTACC 80 CGTCAGCCTATCCCCAAGGCACCACCGGCCAGGAGGAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACGCCGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACGCCGCGGCAGGGCAG
SEO ID NO: 145 143 144 140 139 142 141 139-145 co SEO ID NO: 145 143 144 140 139 142 141 139-145 co	ISOLATE  DK13 1 26 1 27 27 28 1 25 1 25 1 27 1 1  INSERSUS  ISOLATE  DK13 2 26 2 27 2 28 2 24 2 25 2 21 2 INSERSUS  INSERS  INSERSUS  INSERSUS  INSERS  IN	84 CGCCAGCCTATCCCCAAGGCgCGCCAACTCGAGGGLAGGTCCTGGGCTCAGCCTGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACCGTCGGTCGAGGGAAGGTCCTGGGCTCAGCCCGGGTACC 85 CGTCAGCCTATCCCCAAGGCACCGTCGGTCGAGGGCAGGTCCTGGGCTCAGCCGGGTACC 86 CGTCAGCCTATCCCCAAGGCACCGCGGTCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 87 CGCCAGCCTATCCCCAAGGCACCGCGTCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 88 CGTCAGCCTATCCCCAAGGCACCGCGTCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 88 CGTCAGCCTATCCCCAAGGCACCGCGTCGAGGGCAGGTCCTGGGCTCAGCCCGGTACC 88 CGTCAGCCTTTACGGAATGAGGCTCGGGTCGGGGGGAGGTCCTGGTCACCCCGTGG 85 CATGGCCCTTTACGGAATGAGGGTTCGGGTTGGGCGGGATGGCTCCTGTCACCCCGTGG 85 CATGGCCTCTTTACGGTAATGAGGGTTCGGGTTGGGTCAGGAGTTGCTCCTTTGCCCCCGGG 85 CATGGCCTCTTTTACGGTAATGAGGGTTCGGGTTCGGGTCAGGAGGTTGCTCCCCCGGG 85 CATGGCCCTTTTACGGTAATGAGGGTTCGGGTTCGGGTCAGGAGGTTGCTCCCCCGGG 85 CTTGGCCCCTTTTACGGAATGAGGGTTCGGGTTCGGGTGCAGGTTGCTCCCCCGGG 85 CTTGGCCCCTTTTAGGCAATGAGGGTTCGGGTTGGGCAGGTTGGCTCCTTTCCCCCGGG 85 CTTGGCCCCTTTTAGGGAATGAGGGTTCTGGGTTGGAGGAGGTTGCTCCTTTCCCCCGGG 85 CTTGGCCCCTTTTAGGGAATGAGGGTTCTGGGTTGGGAGGAGGTTGCTCCTTTTCCCCCCGGG 85 CTTGGCCCCTTTTAGGGAATGAGGGTTTGGGTTGGGTGGAGGTTGCTCCTTTCCCCCCGGG 85 CTTGGCCCCTTTTAGGGAATGAGGGTTTGGGTTGGGTGGAGGTTGCTCCTTTCCCCCCGGG 85 CTTGGCCCCTTTTAGGGAATGAGGGTTTTGGGTTGGGAGGAGGTTGGCTCCTTTTCCCCCCGGG CCTTGGCCCCTTTTAGGGAATGAGGGTTTTGGGTTGGG
SEO ID NO: 145 143 144 140 139 141 139-145 CO SEO ID NO: 145 144 140 139 144 141 139-145 CO	ISOLATE  DK13 1 26 1 27 2 28 1 25 1 25 1 27 1 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 28	84 CGCCAGCCTATCCCCAAGGCGCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCGCCGAACLCGAGGAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTACC 85 CGTCAGCCTATCCCCAAGGCACGTCGGGTCGAGGGAGGTCCTGGGCTCAGCCGGGTACC 86 CGCCAGCCTATCCCCCAAGGCACGCGGTCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 87 CGCCAGCCTATCCCCAAGGCGCCCGGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 88 CGCCAGCCTATCCCCAAGGCGCCCGGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCGCCCGGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC CGCCAGCCTATCCCCAAGGCACCCGGGTCGAGGGCAGGTCCTGGCTCAGCCCCGTGG 85 CATGGCCCTTTACGGAATGAGGGTTCCGGGTGGGCGGAATGGCTCCTGTCACCCCGTGG 85 CATGGCCTCTTTACGGTAATGAGGGTTTCGGGTGGGCAGGATGGCTCCTGTCACCCCGTGG 85 CATGGCCCTTTTACGGTAATGAGGGCTTTCGGGTGGGCAGGTTGGCTCCTGTCCCCCGGGG 85 CTTGGCCCCTCTATAGGTAATGAGGGCTTGCGGTGGGCAGGGTTGGCTCCTGTCCCCCCGGG 85 CTTGGCCCCTTTATGGCAATGAGGGCTTTGGGTGGGCAGGGTTGGCTCTGTCCCCCCGGG 85 CTTGGCCCCTTTATGGCAATGAGGGCTTTGGGTGGGCAGGGTTGGCTCTGTCCCCCCGGG CCTTGGCCCCTTTATGGCAATGAGGGCTTTGGGTGGGCAGGGTTGGCTCTGTCCCCCCGGG CCTTGGCCCCTTTATGGCAATGAGGGCTTGGGTGGGCAGGGTTGGCTCTGTCCCCCCGGG CCTTGGCCCCTTTATGGCAATGAGGGCTTGGGTGGGCAGGGTTGGCTCCTGTCCCCCCGGG CCTTGGCCCCTTTATGGCAATGAGGGCTTGGGTGGGCAGGGTTGGCTCCTGTCCCCCCGGG CCTTGGCCCCTTTATGGCAATGAGGGCTTTGGGTTGGG
SEO ID NO: 145 143 144 140 139 142 141 139-145 co SEO ID NO: 141 139-145 co SEO ID NO: 141 139-145 co	ISOLATE  DK13 1 26 1 27 1 28 1 25 1 25 1 27 1 1 INSERSUS  ISOLATE  DK13 2 26 2 27 2 28 2 21 2 21 2 INSERSUS  ISOLATE  DK13 3 26 3	84 CGCCAGCCTATCCCCAAGGCgCGCCAACTCGAGGGAAGGTCCTGGGCTCAGCCTGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 84 CGTCAGCCTATCCCCAAGGCACCGTCGGTCGAAGGGAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACCGTCGGTCGAAGGGAGGTCCTGGGCTCAGCCGGGTACC 85 CGTCAGCCTATCCCCAAGGCACCGCGGTCGAAGGGCAGGTCCTGGGCTCAGCCGGGTACC  CGCCAGCCTATCCCCAAGGCACCGCGGTCGAAGGGAAGG
SEO ID NO: 145 143 144 140 139 142 141 139-145 co  SEO ID NO: 145 140 139 142 141 139-145 co	ISOLATE  DK13 1 26 1 27 2 28 1 25 1 25 1 27 1 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 3  ISOLATE  DK13 3 27 3	84 CGCCAGCCTATCCCCAAGGCGCCCAACLCGAGGGLAGGTCCTGGGCTCAGCCLGGGTATC 84 CGCCAGCCTATCCCCAAGGCGCCGAACLCGAGGGAGGTCCTGGGCTCAGCCCGGGTATC 85 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAAGGTCCTGGGCTCAGCCGGGTATC 86 CGCCAGCCTATCCCCAAGGCACGTCGATCTGAGGGAAAGGTCCTGGGCTCAGCCGGGTACC 86 CGTCAGCCTATCCCCAAGGCACGTCGGTCCGAGGGAAGGTCCTGGGCTCAGCCCGGGTACC 86 CGTCAGCCTATCCCCAAGGCACCGGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 87 CGCCAGCCTATCCCCAAGGCGCCCGGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 88 CGCCAGCCTATCCCCAAGGCGCCCGGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 89 CGCCAGCCTATCCCCAAGGCGCCCGGGTCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACC 80 CCCAGCCTATCCCCAAGGCACCCGGGTCGAGGGGGGGGGG
SEO ID NO: 145 143 144 140 139 142 141 139-145 co SEO ID NO: 141 139-145 143 144 140 139-145 141 139-145 141 139-145 141 139-145 141 145 146 147 148 149 149 149 149 149 149 149 149 149 149	ISOLATE  DK13 1 26 1 27 2 28 1 25 1 25 1 27 1 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 2 28 2 27 3  ISOLATE  DK13 3 27 3	84 CGCCAGCCTATCCCCAAGGCgCGCCAACTCGAGGGAAGGTCCTGGGCTCAGCCTGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 84 CGCCAGCCTATCCCCAAGGCACCGTCGATCTGAGGGAAGGTCCTGGGCTCAGCCGGGTATC 84 CGTCAGCCTATCCCCAAGGCACCGTCGGTCGAAGGGAGGTCCTGGGCTCAGCCGGGTACC 84 CGTCAGCCTATCCCCAAGGCACCGTCGGTCGAAGGGAGGTCCTGGGCTCAGCCGGGTACC 85 CGTCAGCCTATCCCCAAGGCACCGCGGTCGAAGGGCAGGTCCTGGGCTCAGCCGGGTACC  CGCCAGCCTATCCCCAAGGCACCGCGGTCGAAGGGAAGG

# FIGURE 6H

139	Z4	306	CTCTCGGCCATCTTGGGGCCCAAATGATCCCCGGCGGAGATCGCGCAATCTGGGTAAGGTC
142	25		ATCTCGGCCATCTTGGGGCCAAAATGATCCCCGGCGTAGGTCCCGCAATCTGGGTAAGGTC
141	Z1	306	tTCcaGGCCgTCTTGGGGCCccAATGATCCCCGGCGTAGGTCCCGLAATCTGGGTAAAGTC
139-145	consensus		cTCtcGqCCgTCTTGGGGcCcaAATGATCCCCGGCGqAGqTCcCGcAAttTGGGTAAqGTC
SEO ID NO			
145	DK13	367	ATCGATACCTAACTTGCGGCTTCGCCGAcCTCATGGGATACATCCCGGTCGTAGGCGCCC
143	26		ATCGATACECTAACTTGCGGETTCGCCGAECTCATGGGATACATCCCGCTCGTAGGCGCCC
144	Z7		ATCGATACCCTAACcTGCGGCTTtGCCGACCTCATGGGATACATCCCGCTCGTAGGCGCCC
140	Z8		ATCGATACCCTcACGTGCGGCTTCGCCGACCTCATGGGATACATCCCGCTCGTGGGCGCCC
139	Z4	367	ATCGATACCCTGACGTGCGGCTTCGCCGACCTCATGGGATACATCCCGaTCGTGGGCGCCC
142	25		ATCGATACCCTGACGTGTGGCTTCGCCGACCTCATGGGATACATTCCGCTCGTcGGCGCCC
141	Z1	307	ATCGATACCCTGACGTGTGGCTTCGCCGACCTCATGGGATACATTCCGCTCGTGGGCGCCC
141	21	367	ATCGATACCCTGACGTGTGGCTTCGCCGACCTCATGGGATACATTCCGCTCGTAGGCGCCC
139-145	consensus		ATCGATACcCT-ACgTGcGGcTTcGCCGAcCTCATGGGATACATcCCGcTCGTaGGCGCCC
			-
SEO ID NO	: ISOLATE		
145	DK13	420	G000000-000000000000000000000000000000
		428	CCGTGGGtGGCGTCGCCAGAGCCCTGGCGCATGGCGTCAGGCtTCTGGAGGACGGGGTCAA
143	Z6	428	CCGTGGGCGCGTCGCCAGGGCCCTGGCACATGGtGTTAGGGCTGTGGAGGACGGGATCAA
144	27	428	CCGTGGCCGCCTCGCCAGGCCCTAGCGCATGGCGTTAGGGCTCTGGAGGACGGGATLAA
140	Z8		CaGTaGGaGGCGTCGCCAGaGCCCTGGCGCATGGCGTCAGGGCTGTGGAGGACGGGATCAA
139	24		
		428	CcGTgGGgGGCGTCGCCAGGGCtCTGGCGCATGGCGTCAGGGCTGTGGAGGACGGGATLAA
142	25	428	CaGTaGGTGGCGTCGCCAGGGCCLTGGCGCATGGCGTCAGGGCCCTGGAGGACGGAATCAA
141	Z1	428	CtgtgGGTGGCGTCGCCAGGGCCcTGGCGCATGGCGTCAGGGCCGTGGAGGACGGAATLAA
			The second secon
139-145	consensus		G-07-00-00-00-00-00-00-00-00-00-00-00-00-
139-143	Combembus		CcGTgGGtGGCGTCGCCAGgGCccTgGCgCATGGcGTcAGGgctgTGGAGGACGGgaTcAA
SEO ID NO	: ISOLATE		
145	DK13	489	TTATGCAACAGGGAATCTTCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCACTgCTcTCG
143	Z6	400	TTATGCAACAGGGAATCTTCCCGGTTGCTCTTTCTCTATCTTCCTCTTTGGCACTTCTTTCG
	Z7		
144		489	TTATGCAACAGGGAACCTTCCCGGTTGCTCTTTLTCTATCTTCCTCTTTGGCACTTCTTTCG
140	Z8		CTATGCAACAGGGAACCTTCCtGGTTGCTCTTTCTCTATCTTCCTCTTGGCACTTCTCTCG
139	Z4	489	CTATGCAACAGGGAATCTTCCGGGTTGCTCTTTCTCTATCTTCCTLTTGGCACTTCTLTCG
142	Z5		CTATGCAACAGGGAATCTTCCTGGTTGCTCCTTtTCTATCTTCCTaCTTGCACTTtTCTCG
141	Z1	489	CTACGCAACAGGGAACCTTCCTGGTTGCTCtTTCTCTATCTTtCTtCTTGCACTTCTCTCG
		107	CIACOCAACAGGGAACCIICCIGGIIGCICCIICCIAICIICCICCIIGCACIICICG
139-145	consensus		cTALGCAACAGGGAALCTTCCcGGTTGCTCLTTcTCTATCTTcCTcLTgGCACTLcTcTCG
SEO ID NO	: ISOLATE		
145	DK13	550	TGCCTgACTGTTCCCgCtTCGGCC
143	Z6	550	TOCOTE TOTAL CONTROL OF
			TGCCTAACTGTTCCCACCTCGGCC
144	Z7		TGCCTgACTGTTCCCGCCTCGGCC
140	Z8	550	TGCCT&ACcGTcCCAGCGTCtGCT
139	Z4		TGCCTcACtGTtCCAGCGTCgGCT
142	25		TGCtTGACAACACCGGCATCcGCT
		220	TOCETON CACCOGCAT COCCT
141	<b>Z</b> 1	550	TGCcTGACAACACCaGCATCtGCc
139-145	consensus		TGCcTgACtgttCC-qC-TCqGCc

# FIGURE 61

SEO ID NO: ISOLATE	
153 SA11	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC&AAAGAAACACCAACCGCCGCCCACAGG
152 SA6	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCCAAAGAAACACCGACCG
	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGCCGCCCACAGG
147 SA5	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGCCCCCCCACAGG
148 SA7	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGCCGCCCACAGG
149 SA1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCtCCGCCCACAGG
150 SA3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGCCGCCCACAGG
151 SA13	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCCAAAGGAAACACCAACCGCCGCCCACAGG
131 3413	1 Aldrochedanie i investicananana communication con control
146 153	100100100100111000111001110011100111001100010010010010001000100010000
146-153 consensus	ATGAGCACGAATCCTAAACCTCAAAGAAAAACC&AAAGAAACACCAACCgCCGCCCACAGG
SEO ID NO: ISOLATE	
153 SA11	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
152 SA6	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
146 SA4	62 ACGTLAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTCTACTTGTTGCCGCGCAGGGG
147 SA5	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
148 SA7	62 ACGTCAAGTTCCCGGGCGGTCGTCAAATCGTTGGTGGAGTTTACTTGTTGCGGCAGGGG
149 SA1	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
150 SA3	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
151 SA13	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGG
146-153 consensus	ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTLTACTTGTTGCCGCGCAGGGG
SEO ID NO: ISOLATE	
153 SA11	102 0000 0 0000000000000000000000000000
	123 CCCTaGgtTGGGTGTGCGCGCGACTCGGAAGACTTCaGAACGGTCGCAACCCCGTGGgCGG
152 SA6	123 CCCTcGtaTGGGTGTGCGCGCGACTCGGAAGACTTCgGAACGGTCGCAACCCCGTGGaCGG
146 SA4	123 CCCTAGgTTGGGTGTGCGCGCGACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGGCGG
147 SA5	123 CCCTAGATTGGGTGTGCGCGCGACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGGCGG
148 SA7	123 CCCTAGGTTGGGTGTGCGCGCGACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGGCGG
149 SA1	123 CCCCAGGTTGGGTGTGCGCGCGACTCGGAAGACTTCgGAACGGTCGCAACCCCGTGGGCGG
150 SA3	123 CCCCAGGTTGGGTGTGCGCGGACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGACGG
151 SA13	123 CCCLAGGTTGGGTGTGCGCGCACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGACGG
IJI SAIJ	123 CCCCAGGIIGGGIGGGCGCAACICGGAAGACIICAGAACGGICGCAACCCCGIGGACGG
146-153 consensus	000 - 0 - magamamagagaga - magaga - 0 - maga
146-153 Consensus	CCCtaGgtTGGGTGTGCGCGCGACTCGGAAGACTTCaGAACGGTCGCAACCCCGTGGgCGG
SEO ID NO: ISOLATE	
153 SA11	184 CGTCAGCCTATTCCCAAGGCGCGCCAAcCCaCGGGcCGGTCCTGGGGTCAACCCGGGTACC
152 SA6	184 CGTCAGCCTATTCCCAAGGCGCGAAtCCgCGGGtCGGTCCTGGGGTCAACCCGGGTACC
146 SA4	184 CGCCAGCCTATTCCCAAGGCGCGCCAACCCACGGGCCGGTCCTGGGGTCAACCCGGGTACC
147 SA5	184 CGCCAGCCTATTCCCAAGGCGCCCAACCCACGGCCGGTCCTGGGGTCAACCCGGGTACC
148 SA7	184 CGCCAGCCTATTCCCAAGGCGCCCAACCCACGGGCCGGTCCTGGGGTCAACCCGGGTACC
149 SA1	184 CGCCAGCCTATTCCCAAGGCGCCCCCCACCCCCCGGCCCGGTCCTGGGGTCAACCCCGGGTACC
150 SA3	184 CGCCAGCCTATTCCCAAGGCLCGCCCACCGGGCCGGTCCTGGGGTCAACCCGGGTACC
151 SA13	184 CGtCAGCCTATeCCCAAGGCgCGCCAGCCCACGGGCCGGTCCTGGGGTCAACCCGGGTACC
146-153 consensus	CGcCAGCCTATtCCCAAGGCgCGCCAacCCaCGGGcCGGTCCTGGGGTCAACCCGGGTACC
SEO ID NO: ISOLATE	
153 SA11	245 CTTGGCCCLTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGCTGCTCTCCCCLCGAGG
152 SA6	245 CTTGGCCCCTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTTGCTCCCCCCCC
146 SA4	
	245 CTTGGCCCCTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTTGCTCTCCCCCCGAGG
147 SA5	245 CTTGGCCCCTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTTGCTCCCCCCGAGG
148 SA7	245 CTTGGCCCCTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTTGCTCTCCCCCCGAGG
149 SA1	245 CTTGGCCCCTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTTGCTCCCCCCGAGG
150 SA3	245 CTTGGCCCCTTTACGCCAATGAGGGCCTCGaGTGGGCAGGGTGGTTGCTCTCCCCCCGAGG
151 SA13	245 CTTGGCCCCTTTAEGCCAATGAGGGCCTCGgGTGGCAGGGTGGTTGCTCTCCCCCCGAGG
146-153 consensus	CTTGGCCCcTTTAcGCCAATGAGGGCCTCGgGTGGCAGGGTGGLTGCTCTCCCCcCGAGG

#### FIGURE 61

SEO ID NO: ISOLATE	
153 SA11	306 CTCTCGGCCTAAcTGGGGCCCCAATGACCCCCGGCGAAGATCGCGCAATTTGGGCAAGGTC
152 SA6	306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCCGGCGAAAATCGCGCAATTTGGGTAAGGTC
146 SA4	306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCCGGCGAAAGTCGCGCAATTTGGGTAAGGTC
147 SA5	306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCCGGCGAAATCGCGCAATTTGGGTAAGGTC
148 SA7	306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCCGGCGAAAGTCGCGCAATTTGGGTAAGGTC
149 SA1	306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCCGGCGGAAGTCGCGCAATTTGGGTAAGGTC
150 SA3	306 CTCTCGGCCTAgTTGGGGCCCCAAcGACCCCCGGCGGAAATCGCGCAATTTGGGTAAGGTC
151 SA13	306 CTCTCGGCCTAaTTGGGGCCCCAALGACCCCCGGCGGAAATCGCGCAACTTGGGTAAGGTC
isi onis	300 DIETEGGETALITOSOCIECANESACCECEGGEGGAARTEGGEGAACTIGGGTAAGGTC
146-153 consensus	CTCTCGGCCTA&tTGGGGCCCCAAtGACCCCCGGCGaA&aTCGCGCAAtTTGGGtAAGGTC
146-133 COMBEMBUB	CICICGGCCIAZ 19900CCCCAACGACCCCCGGCGAAZ 1CGCGCAACIIGGGCAACGIC
SEO ID NO: ISOLATE	
	3.62 Amogama agama a
153 SA11	367 ATCGATACCCTAACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTAGGCGGCC
152 SA6	367 ATCGATACCCTAACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTAGGCGGCC
146 SA4	367 ATCGATACCCTAACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTAGGCGGCC
147 SA5	367 ATCGATACCCTAACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTAGGCGGCC
148 SA7	367 ATCGACACCCTAACATGCGGATTCGCCGACCTCATGGGGGTACATCCCGCTCGTAGGCGGCC
149 SA1	367 ATCGATACCCTAACGTGCGGATTCGCCGACCTCATGGGGGTACATCCCGCTCGTAGGCGGCC
150 SA3	367 ATCGATACCCTAACGTGCGGATTCGCCGAtCTCATGGGGTACATCCCGCTCGTAGGCGGCC
151 SA13	367 ATCGATACCCTGACGTGCGGATTCGCCGAcCTCATGGGGTACATCCCGCTCGTAGGCGGCC
146-153 consensus	ATCGAŁACCCTAACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTAGGCGGCC
SEO ID NO: ISOLATE	
153 SA11	428 CCGTTGGGGGCGTCGCAAGGGCCCTCGCACACGGTGTGAGAGCTCTTGAGGACGGGGTAAA
152 SA6	428 CCGTTGGGGGCGTCGCAAGGGCLCTCGCACACGGTGTGAGGGTTCTTGAGGACGGGGTAAA
146 SA4	428 CCGTTGGGGGCGTCGCAAGGGCCCTtGCACATGGTGTGAGGGTTCTTGAGGACGGGGTAAA
147 SA5	428 CCGTTGGGGGCGTCGCAAGGGCCCTCGCACATGGTGTGAGGGTTCTTGAGGACGGGGTAAA
148 SA7	428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGTTCTTGAGGACGGGGTAAA
149 SA1	428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGTTCTTGAGGACGGGTAAA
150 SA3	428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACALGGTGTGAGGGTTCTTGAGGACGGGGTAAA
151 SA13	428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGGTCCTTGAGGACGGGGTAAA
	Collocation Co
146-153 consensus	CCGTTGGGGGCGTCGCAAGGGCtCTcGCACAcGGTGTGAGgGttCTTGAGGACGGGGTAAA
	CCCTTCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCC
SEO ID NO: ISOLATE	
153 SA11	489 tTATGCAACAGGGAATCTtCCCGGTTGCTCTTTCTCcATCTTTATCCTTGCACTTCTCTCG
152 SA6	489 CTATGCAACAGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTTTCTCTTGCACTTCTCTCG
146 SA4	
140 SA4	
148 SA7	489 CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG 489 LTACGCAACAGGGAATCTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG
149 SA1	
150 SA3	489 CTACGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTTTCC 489 CTACGCAACAGGGAATTTACCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTTTCA
151 SA13	
131 SA13	489 CTALGCAACAGGGAATTTACCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTTTCA
146-153 consensus	-margana - aggana - a
146-153 consensus	CTAtGCAACaGGGAATtTgCCCGGTTGCTCTTTCTCtATCTTTaTCCTTGCACTTCTCTCg
CEO TO NO. TOOLSED	
SEO ID NO: ISOLATE	FF0
153 SA11	550 TGCtTgACCGTCCCgGCCaCTGCA
152 SA6	550 TGCCTaACCGTCCCtGCCTCTGCA
146 SA4	550 TGCCTGACCGTCCCgGCCTCTGCA
147 SA5	550 TGCETGACCGTCCCAGCCTCTGCA
148 SA7	550 TGCCTGACCGTCCCAGCCTCcGCA
149 SA1	550 TGECTGAECATCCCGGCCTCTGCA
150 SA3	550 TGCCTGACCGTCCCGGCCTCTGCA
151 SA13	550 TGCCTGACLGTCCCGACCTCTGCc
146-153 consensus	
146-155 Consensus	TGccTgAccgTCCCggCCtCtGCa

#### FIGURE 6J

SEO ID NO: Gen	otype		
	cons.	1 2	ATGAGCACgaaTCCtAAACCtCAAAGAaAaACCaaAcGtAACACcAaCCgcCGCCCacagG
103-124	1	1	ATGAGCACGA&TCCTAAACCTCAAAGA&A&AACCAAACGTAACACCA&CCGcCGCCCACAGG
125-134	2		ATGAGCACAAATCCEAAACCECAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCACAGG
135-138	3		ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAAGAAACACCATCCGTCGCCCACAGG
139-145	4		ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCGACCG
146-153	5		ATGAGCACGAATCCTAAACCTCAAAGAAAAACCaAAAGAAACACCAACCGCCCCACAGG
154	6		ATGAGCACACTTCCAAAACCCCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCAACGG
134	•		AT GAGCACACT T CCAMARCCC CAMAGAMAMACCAMACCAMCCG T CGCCCAMCGG
			• •
SEO ID NO: Gen			·
103-154	cons.	62	AcgTcAAgTTcCCgGGcGGtGGtCAGATCGTtGGtGGAGTtTActTGtTGCCGCGCAGGGG
103-124	1	62	ACGTCAAGTTCCCGGGCGGCGGCCAGGTCGTCGTGGAGTLTAccTGTTGCCGCGCAGGGG
125-134	2	62	ACGTLAAGTTcCCGGGcGGGCCAGATCGTTGGCGGAGTaTACTTGcTGCCGCGCAGGGG
135-138	3	62	ACTTCAAGTTCCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG
139-145	4	62	AcGTaAAgTTCCCgGGtGGtGGcCAGATCGTTGGCGGAGTTTACTTGtTGCCGCGCAGGG
146-153	5	62	ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTLTACTTGTTGCCGCGCAGGGG
154	6	62	ACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGCGGAGTTTACTTGTTGCCGCGCAGGGG
	•		The state of the s
SEO ID NO: Gen	otime		the state of the s
	cons.	123	CCC and Cat TCCC TCTCCCC accorded and an accordance and accordance
103-154	COMB.	123	CCCcaGgtTGGGTGTGCGCgCgaCtaGgAAgaCTTCcGAgCGgTCgCAaCCtcGtGGaaGg
102 124		100	: 000 - 0 - mm000m0m000000 - 0 - 1001101 mm0 - 01000 - 0 - 1001101
103-124	1	123	CCCcaGgTTGGGTGTGCGCGCgaCtAGGAAGACTTCcGAGCGgTCgCAACCTCGtGGaaGg
125-134	2	123	CCCcAGgTTGGGTGTGCGCGCGACAAGGAAgaCTTCcGAgCGaTCCCAGCCGCGTGGgAGg
135-138	3	123	CCCACGATTGGGTGTGCGCGCGACGCGTAAAACTTCTGAACGGTCaCAGCCTCGCGGACGa
139-145	4	123	CCCcaGgTTGGGTGTGCGCgCgaCTcGgAAGACTTCGGAGCGGTCgCAACCTCGtGGcAGg
146-153	5	123	CCCtaGgtTGGGTGTGCGCGGACTCGGAAGACTTCaGAACGGTCGCAACCCCGTGGgCGG
154	6	123	CCCCGGTTGGGTGTGCGCGCGACGAGAAAGACTTCCGAGCGATCCCAGCCCAGAGGCAGG
SEO ID NO: Gen	otvoe		
	cons.	184	CGaCAgCCtATcCCcaAgGctCGcCggcccgagGGcaggtcCTGGGctcagCCcGGgtAcC
103-124	1	184	CGaCAaCCTATCCCCAAGGCtCGcCggCCCGAGGGCAGGGCCTGGGCtCAGCCcGGGtAcC
125-134	2	184	
	3		CGCCAGCCCATCCCgAAAGATCGGCGCtCCACtGGCAAGtCCTGGGGAAaaCCaGGATAtC
135-138		184	CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGGTCCTGGGCTCAGCCcGGGTACC
139-145	4	184	CGtCAgCCTATCCCCaAGGCaCGtCggtccGAGGGCAGgTCCTGGGCtCAgCCcGGGTAcC
146-153	5	184	CG-CAGCCTATtCCCAAGGCgCGCCAa-CCaCGGG-CGGTCCTGGGGTCAACCCGGGTACC
154	6	184	CGCCAACCTATACCAAAGGCGCGCCAGCCCCAGGGCAGGCA
SEQ ID NO: Gene			
103-154	cons.	245	CtTGGCCccTcTAtGgcaAtGAgGGcttcGggTGGGCaGGaTGGcTccTgTCcCCcCgcGG
103-124	1	245	CtTGGCCCCTCTAtGgCaAtGAGGGCttgGGgTGGGCaGGATGGCTCCTGTCaCCCCgtGG
125-134	2	245	CtTGGCCcTgTAtGGgAAtGAGGGctCGGCTGGCCAGGtTGGCTCCTGTCCCCCCGGGG
135-138	3	245	CTTGGCCCCTCTATGGTAAcGAGGGCTGCGGGTGGGCAGGgTGGCTCCTGTCCCCACGCGG
139-145	4	245	CtTGGCCtCTtTAcGGcAAtGAgGGcTGcGGGTGGGCAGGGTGGCTCcTGTCcCCcCGcGG
146-153	5	245	
154	6		CTTGGCCCCTTTAcGCCAATGAGGGCCTCGgGTGGGCAGGGTGGLTGCTCCCCcCGAGG
154	6	245	CTTGGCCTCTTTATGGAAACGAGGGCTGTGGGTGGGCAGGTTGGCTCCTGTCCCCCCGCGG
CEO TO NO. C			
SEO ID NO: Gene			
103-153	cons.	306	cTCtcggCCtagtTGGGGcCccActGAcCCCCGGCgtaggTCgCGcAAttTGGGtAagGTC
103-124	1	306	cTCtCGGCCTAgtTGGGGCCCcAcaGACCCCCGGCGtAGGTCGCGtAATtTGGGtAAggTC
125-134	2	306	tTCtCgtCCttctTGGGGCCCCActGAcCCCCGGCAtAgaTCgCGcAActTGGGtAaqGTC
135-138	3	306	CTCCCGTCCATCTTGGGGCCCAAAcGACCCCCGGCGgaGGTCCCGCAATTTGGGTAAAGTC
139-145	4	306	CTCLcGgCCgTCTTGGGGcCcaAATGATCCCCGGCGGAGgTCcCGcAALLTGGGTAAGGTC
146-153	5	306	CTCTCGGCCTAatTGGGGCCCCAAtGACCCCCGGCGAAaaTCGCGCAAtTTGGGtAAGGTC
154	6	306	CTCCCGGCCACATTGGGGCCCCAATGACCCCCGGCGTCGATCCCGGAATTTGGGTAAGGTC
	-		The state of the s

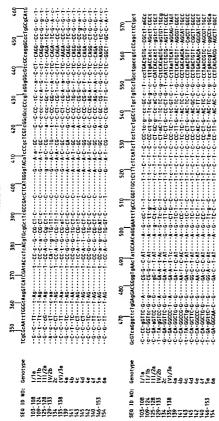
# FIGURE 6J

SEO ID NO:	Genotype		• 7
103-154	cons.	367	ATCGALACccTcACgTGcgGctTcGCCGAcCTCATGGGgTACaTcCCgcTCGTcGGcGccC
103-124	1	367	ATCGALACCCTcACaTGCGGCTTcGCCGACCTCATGGGGTACATLCCGCTCGTCGGcGccC
125-134	2	367	ATCGALACCCTAACGTGcgGttTTGCCGACCTCATGGGGTACaTcCCcGTCGTtGGCGccC
135-138	3	367	ATCGATACCCTLACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC
139-145	4	367	ATCGATACCTGACGTGCGGCTTCGCCGACCTCATGGGATACATCCCGCTCGTAGGCGCCC
146-153	5	367	ATCGAŁACCCTAACGTGCGGATTCGCCGACCTCATGGGGTACATCCCGCTCGTAGGCGGCC
154	6	367	ATCGATACCCTAACGTGTGGGTTCGCCGATCTCATGGGGTACATTCCCGTCGTGGGCGCGC
SEQ ID NO:	Genotype		
103-154	cons.	428	CcgTaGGgGGCGtcGCcaggGCccTgGCgCAtGGcGTcaGggttcTgGAqGACGGgqTqAA
103-124	1	428	CccTaGGgGGcGcTGCCAGgGCccTGGCqCAtGGcGTCCGgGTtcTGGAqGACGGCGTGAA
125-134	2	428	CggTtGGaGGcGTcGCCAGAGCtCTqGCaCAtGGtGTqAGqGTcCTGGAGGACGGgaTaAA
135-138	3	428	CCGTAGGAGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
139-145	4	428	CcGTgGGtGGCGTCGCCAGgGCccTgGCgCATGGcGTcAGGgctgTGGAGGACGGgaTcAA
146-153	5	428	CCGTTGGGGGCGTCGCAAGGGCLCTGGCACACGGTGTGAGGGLLCTTGAGGACGGGGTAAA
154	6	428	CTTTGGGCGGCGCGCGGCTGCGCTCGCACATGGCGTGAGGGCAATCGAGGACGGATCAA
SEO ID NO:	Genotype		
103-154	cons.	489	cTatGCAACaGGGAAttTgCCcGGTTGCtCtTTcTCtATcTTccTccTgGCtcTqcTqTCc
103-124	1	489	cTAtGCAACAGGGAAtcTqCCcGGTTGCtCtTTcTCTATCTTCCTctTqGCttTGcTqTCc
125-134	2	489	tTAtGCAACaGGqAAttTqCCtGGTTGCTCtTTtTCTATcTTctTqcTtGCccTtcTGTCc
135-138	3	489	TTTcGCAACAGGGAACTTGCCCGGTTGCTCcTTTTCTATCTTCCTTCTTGCcCTGTTCTCL
139-145	4	489	CTAtGCAACAGGGAAtCTTCCcGGTTGCTCtTTCTCTATCTTCCTctTgCACTtcTCTCG
146-153	5	489	CTALGCAACAGGGAATLTGCCCGGTTGCTCTTTCTCLATCTTTATCCTTGCACTTCTCTCT
154	6	489	TTATGCAACAGGGAATCTCCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCACTACTCTCG
SEO ID NO:	Genotype		
103-154	cons.	550	TGcctgaccgtcCCagcttCtqct
103-124	1	550	TGttTgACcatcCCaGctTCcGCt
125-134	2	550	TGCatCaCagtgCCaGtgTCtGCt
135-138	3	550	TGCtTAaTTCATCCaGCAGCTAGT
139-145	4	550	TGCcTgACtgttCCagCgTCgGCc
146-153	5	550	TGccTgAccgTCCCggCCtCtGCa
154	6	550	TGCCTCACAACGCCAGCTTCGGCT

# FIGURE 6K

	MICHICAGONIC COMMINICAMA CANADA CANCER CONTROL	230	10.00	24 25 26 20 270 280 280 300 310 310 320 330 340 creatic conference of co	C
=		129 130 140 150 160 170 189 190 200 210 220 233 Galadaceccesaga tracelasegas tracelase dages consequentes establistas establistas establistas establistas esta	\$6555555555555555555555555555555555555	340	20100000000000000000000000000000000000
	500000000000000000000000000000000000000	220	100 CT - CT	3333	•••••
<u>6</u> _		6980	00000000000000000000000000000000000000	330 Acto	0.000000000000000000000000000000000000
	9	210	AAA C C C C C C C C C C C C C C C C C C	ິ່ງ	000000000000000000000000000000000000000
8-	CAGATGAT	1000		320	
•	G · · · · · · · · · · · · · · · · · · ·	ور اور	004440000000000	3	CATCTTTTTTT
-	9 0100101010101010	002	333323333333333	2066	1000-14 1000-14 1000-14 1000-14 1000-14 1000-14 1000-14 1000-14 1000-14 1000-14 1000-14
8-	8	CtATo		310 -10tc	- + + + + + + + + + + + + + + + + + + +
	5 0000000000000000000000000000000000000	061 De 42	0.000004000004	9936	5 86 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
2-	5 000000000000000000000000000000000000	90069	400 a0 a00 400 600	<u>ي</u> 1	
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	88 88	<b>186552555555</b>	19TC	99999999999999
09	ACALCTRIC COCCUS cog Side grand (ACALCTRIC COCCUS COCUS COCCUS CO	teg	**************************************	23 Ec. C	
	# 4444400040F44	170 ——gCAac		68	444F0B00440F0F
S	ACCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	t ggTCg		280	4
	CCABCCCCABCCCCCABCCCCCCCCCCCCCCCCCCCCC	<u>∑</u>	****	24	550000000000000000000
	¥ dddddddddddddd	160		ot tee	66-91-91-91-91-91-91-91-91-91-91-91-91-91-
4-	MCC 8 8 C G C C C C C C C C C C C C C C C	_ ¥	99994499999999	270 AgGGc1	
	MCC88ACGU	\$ <b>g</b>	1.00 - 0.	.a¥t6	111010101111111
£	CVYCY	<b>e</b> 6ე6;	20 - 20 - 20 - 20 - 20 - 20 - 20 - 20 -	% - \$	
	CAAA CAA CAA CAA CAA CAA CAA CAA CAA CA	5,75		50	000000000000000000000000000000000000000
2-	AAACÉTICAAGA	9000		230	
	3	130 Cgt		500 t1	
5-	Ggaarccaaactro	8000	4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	240 GGgtA	
	MGAGGAG991CCC MAN	129 AGGGG	76222222222222222222222222222222222222	7 000	TOAG - C - G - C - C - C - C - C - C - C -
	¥ ::::::::::::::::::::::::::::::::::::	- 8		Co	22444222222222
	*	*		•	
	Genotype 1/1a 11/1b 11/2b 2c (V)/3a 4b 4b 4c 4d 4d 4d 4d 4d 4d 4d 4d 4d 4d 4d 4d 4d	Genotype	18 7.1b 7.2b 7.2b	Genotype	11,738 736 736
			25 44 45 45 45 45 45 45 45 45 45 45 45 45		22528843494588
	SEG 1D NO: 108-108-108-109-124-128-128-133-138-138-138-138-138-138-138-138-13	SEQ ID NO:	128 133 133 153	SEG ID NO:	38 33
	S 50505755575555555555555555555555555555	SEO	103-108 109-124 125-128 134-138 139-139 143 143 145 146-153	SE0 1	103-108 109-124 125-128 135-138 135-138 141 141 142 143 145 146-153





#### FIGURE 7A

SEO ID NO		
156	US11	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
157	S14	1 MSTNPKPORKTKRNTNRRPODVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR
158	SW1	1 MSTNPKPORKTKRNTNRRPODVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR
159	S18	1 MSTNPKPORKTKRNTNRRPODVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR
160	DR4	1 MSTNPKPORKTKRNTNRRPODVKFPGGGOIVGGVYLLPRRGPRIGVRATPKTGERGÕDDGD
155	DK7	1 MSTNPKPORKTKRNTNRRPODVKFPGGGOIVGGVYLLPRRGPRLGVRADRKTSERSOPRGR
		• •
155-160	consensus	MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAtRKTSERSOPRGR
SEO ID NO:		
156	US11	62 ROPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
157	S14	62 ROPI PKARRPEGRTWAOPGY PWPLYGNEGCGWAGWILLS PROSR PSWGPTD DDD DDD SDNI GRU
158	SW1	62 ROPIPKARRPEGRTWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDDDDDSDNIGKV
159	S18	62 ROPIPKARRPEGRTWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNIGKV
160	DR4	62 ROPIPKARRPEGRTWAOPGYPWPLYGNEGCGWAGWT.LSPRGSPPSWGPTDPPDDSDNT.GKU
155	DK7	62 ROPIPKARRPEGRTWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
155-160	consensus	RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
SEO ID NO:		
156	US11	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
157	S14	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
158	SW1	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
159	S18	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
160	DR4	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
155	DK7	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
155 160		
155-160	consensus	IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
SEO ID NO:	ISOLATE	
156	US11	104 Gramman
157	S14	184 CLTVPASA
158	SW1	184 CLTVPASA
159	SNI S18	184 CLTVPASA 184 CLTVPASA
160	DR4	
155	DK4 DK7	184 CLTVPASA
133	DK/	184 CLTVPASA
155-160	consensus	CLTVPASA
122-160	Compensus	CLIVPASA

# FIGURE 7B

	72
SEQ ID NO: ISOLAT   175	1 MSTLPKPORKTKRNTSRFPODVKFPGGGLIVGGVYLLPRRGPRLGVRATTKTSRESOPRGR 1 MSTNPKPÖRKTKRNTNRFPODVKFPGGGLIVGGVYLLPRRGPRLGVRATTKTSRESOPRGR 5 1 MSTNPKPÖRGTKRNTNRFPÖDVKFPGGGLIVGGVYLLPRRGPRLGVRATTKTSRESOPRGR 6 1 MSTNPKPÖRGTKRNTNRFPÖDVKFPGGGLIVGGVYLLPRRGPRLGVRATTKTSRESOPRGR 7 MSTNPKPÖRKTKRNTNRFPÖDVKFPGGGLIVGGVYLLPRRGPRLGVRATTKTSRESOPRGR 8 MSTNPKPÖRKTKRNTNRFPÖDVKFPGGGLIVGGVYLLPRRGPRLGVRATTKTSRESOPRGR 8 MSTNPKPÖRKTKRNTNRFPÖDVKFPGGGLIVGGVYLLPRRGPRLGVRATTKTSRESOPRGR 9 MSTNPKPÖRKTKRNTNRFPÖDVKFPGGGLIVGGVYLLPRRGPRLGVRATTKTSRESOPRGR 1 MSTNPKPÖRKTKRNTNRFPÖDVKFPGGGLIVGGVYLLPRRGPLIGVRATTKTSRESOPRGR
161-176 consensu	B MSTnPKPQRkTKRNTnRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR
SEO ID NO: ISOLAT   175	62 ROPIPKARRPEGRAMAQPGHPPLYJANEGIGNAGKLISPRGSRPSWGPTDPRRSSNIGKV 62 ROPIPKARRPEGRAMAQPGHPPLYJANEGIGNAGKLISPRGSRPSWGPTDPRRSSNIGKV 62 ROPIPKARRPEGRAMAQPGHPPLYJGNEGIGNAGKLISPRGSRPSWGPTDPRRSSNIGKV 62 ROPIPKARPEGRAMAQPGYPPLYJGNEGIGNAGKLISPRGSRPSWGPTDPRRSSNIGKV 63 ROPIPKARPEGRAMAQPGYPPLYJGNEGIGNAGKLISPRGSRPSWGPTDPRRSSNIGKV 64 ROPIPKARRPEGRAMAQPGYPPLYJGNEGIGNAGKLISPRGSRPSWGPTDPRRSSNIGKV 65 ROPIPKARRPEGRAMAQPGYPPLYJGNEGIGNAGKLISPRGSRPSWGPTDPRRSSNIGKV 66 ROPIPKARRPEGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 67 ROPIPKARRPEGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 68 ROPIPKARRPEGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 69 ROPIPKARRPEGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 60 ROPIPKARRPEGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 60 ROPIPKARPEGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 60 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 60 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 60 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 60 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 60 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 61 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 62 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 63 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 64 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 65 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSRPSWGPTDPRRSSNIGKV 65 ROPIPKARQPGGRAMAQPGYPPLYJGNEGWAGKLISPRGSSPBWGPTDPRRSSNIGKV 66 ROPIPKARQPGGRAMAGHAGKLISPRGSSPBWGPTDPRRSSNIGKV 67 ROPIPKARQPGGRAMAGHAGKLISPRGSSPBWGPTDPRRSSNIGKV
161-176 consensu	RQPIPKARrPEGRaWAQPGyPWPLYgnEG-GWAGWLLSPrGSRPsWGPtDPRRRSRNLGKV
SEO ID NO:   ISOLAT   175   P	123 IDTLTCGFADLMSY IPLUGGPLGGVARALAHGVKVEDGVNYATGNLPGGSSTSIFLIALLS 124 IDTLTCGFADLMSY IPLUGGPLGGVARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 125 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 126 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 127 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 128 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 129 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 120 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 121 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 122 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 123 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 124 IDTLTCGFADLMSY IPLUGAPLGGAARALAHGVKVLEDGVNYATGNLPGGSSTSIFLIALLS 125 IDTLTCGFADLMSY IPLUGAPLGGVARAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 126 IDTLTCGFADLMSY IPLUGAPLGGVARAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 127 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 128 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 129 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 120 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 121 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 122 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 123 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNYATGNLPGGSFSTFLIALLS 124 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNYATGNLPGGSSTFLIALLS 125 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNATGNURGGSSTFLIALLS 126 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNATGNURGSSTFLIALLS 127 IDTLTCGFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNATGNURGSSTFLIALLS 128 IDTLTCCFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNATGNURGGSSTFLIALLS 129 IDTLTCCFADLMSY IPLUGAPLGGVARALAHGVKVLEDGVNATGNURGGSSTFLIALLS 120 IDTLTCCFADLMSY IPLUGAPLAGRAPHALAHGVKVLEDGVNATGNURGGSSTFLIALLS 121 IDTLTCCFADLMSY IPLUGAPLAGRAPHALAHGVKVLEDGVNATGNURGGSSTFLIALLS
161-176 consensus	IDTLTCGFADLMGYIPLVGaPLGGaARALAHGVRV1EDGVNYATGN1PGCsFSIFLLALLS
SEO ID NO:         ISOLATI           175         Pi           170         IND           162         S45	184 CLTipasa 184 CLTvpasa

84892\_1

171 163 165 169 164 166 167 168 161	S9 D1 P10 IND3 US6 DK1 T10 SW2 SA10 HK4	184 CLTIPASA 184 CLTIPASA 184 CLTIPASA 184 CLTIPASA 184 CLTIPASA 184 CLTIPASA 184 CLTIPASA 184 CLTIPASA 184 CLTIPASA 184 CLTIPASA	
172 176 173	HK3 T3 HK5	184 CLTtPASA 184 CLTiPASA 184 CLTtPVSA	
161-176	consensus	CLTiPaSA	

# FIGURE 7C

SEO ID NO: ISOLATE 173 HK5	
173 HK5	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
176 T3	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
172 HK3	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
174 HK4	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
161 SA10	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
168 SW2 167 T10	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
167 T10 166 DK1	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
164 US6	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
169 IND3	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
165 P10	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
163 D1	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
156 US11	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
157 S14	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
158 SW1	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
159 <u>\$1</u> 8	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR 1 MSTNPKPQRKTKRNTNRRPÖDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
160 DR4	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
155 DK7 170 IND8	1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRAPRKTSERSÖPRGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR 1 MSTNPKPORGTKRNTNRRPODVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR
162 S45 171 S9	1 MSTNPKPORGTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR 1 MSTNPKPORKTKRNTNRRPÖDVKFPGGGÖIVGGVYLLPRRGPRLGVRATRKTSERSÖPRGR
175 P8	1 MSTIPKPORKTIKRITMERPODVEFPGGGOIVGGVILLPREGPRIGVRATEKTSERSOPEGE 1 MSTIPKPORKTERNTSERPODVEFPGGGOIVGGVYLLPREGPRIGVRATEKTSERSOPEGE
1/5	I MSICPAPQAATAANIBAAFQDVAFPGGGQIVGGVIBBPAAGPALGVAATAATSEASQPAGA
155-176 consensus	MSTnPKPQRkTKRNTnRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAtRKTSERSQPRGR
133 1.0 00115011505	TETTI TO QUELLIGIT MATERIAL POR LOCALITATION TO A PROPERTY OF THE PROPERTY OF
SEO ID NO: ISOLATE	
173 HK5	62 RQPIPKARRPEGRtWAQPGYPWPLYGnEGMGWAGWLLSPhGSRPsWGPTDPRRRSRNLGKV
176 T3	62 RQPIPKARRPEGRAWAQPGYPWPLYGdEGMGWAGWLLSPRGSRPNWGPTDPRRRSRNLGKV
172 HK3	62 RQPIPKARQPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRGSRPNWGPTDPRRRSRNLGKV
174 HK4	62 RQPIPKARQPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
161 SA10	62 ROPIPKAROPEGRTWAOPGYPWPLYGNEG1GWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
168 SW2	62 ROPIPKARÓPEGRAWAÓPGYPWPLYGNEGMGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
167 T10 166 DK1	62 ROPIPKAROPEGRAWAOPGYPWPLYGNEGMGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
166 DK1 164 US6	62 RÖPI PKARRPEGRAWAÖPGY PWPLYGNEGMGWAGWLLSPRGSRPSWGPnDPRRRSRNLGKV 62 RÖPI PKARRPEGRAWAÖPGY PWPLYGNEGMGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
169 IND3	62 RÖPIPKARRPEGRAWAÖPGYPWPLYGNEGMGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV 62 RÖPIPKARRPEGRAWAÖPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
165 P10	62 ROPIPKARRPEGRAWAOPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
163 D1	62 ROPI PKARRPEGRAWAOPGY PWPLYGNEGIGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
156 US11	62 ROPIPKARRPEGRTWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNIGKV
157 S14	62 ROPIPKARRPEGRTWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNIGKV
158 SW1	62 RÖPIPKARRPEGRTWAÖPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
159 S18	62 RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
160 DR4	62 RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
155 DK7	62 ROPIPKARRPEGRTWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
170 IND8	62 RÖPIPKARRPEGRAWAÖPGHPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
162 S45	62 ROPIPKARRPEGRAWAOPGHPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV
171 S9 175 P8	62 RÖPIPKARhPEGRAWAÖPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPnDPRRRSRNLGKV
175 P8	62 RQPIPKAR:PEGRAWAQPGhPWPLYaNEGLGWAGWLLSPRGSRPSWGPtDPRRRSRNLGKV
155-176 consensus	ROPIPKAR:PEGRAWAOPGyPWPLYqmEG-GWAGWLLSP:GSRPsWGPtDPRRRSRNLGKV
SEO ID NO: ISOLATE 173 HK5	
173 HK5	123 IDTLTCGFADLMGYIPLVGAPLGGVARALAHGVRVLEDGVNYATGNiPGCSFSIFLLALLS
176 T3	123 IDTLTCGFADLMGYIPLVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
172 HK3	123 IDTLTCGFADLMGYIPLVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
174 HK4	123 IDTLTCGFADLMGYIPLVGAPLGGVARALAHGVRVvEDGVNYATGNLPGCSFSIFLLALLS
161 SA10	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCPFSIFLLALLS
168 Sw2	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
167 T10 166 DK1	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
166 DK1 164 US6	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
169 IND3	
165 P10	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
163 D1	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
156 US11	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
157 S14	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
158 SW1	123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
2	

84892\_1

155-176

#### FIGURE 7C

159	S18	123	IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLIALLS
160	DR4	123	
155	DK7	123	
170	IND8	123	
162	S45	123	
171	S9	123	
175	P8	123	
SEO ID NO: 173 176 176 172 174 1618 168 169 166 164 169 165 163 165 163 165 163 165 165 165 165 165 165 165 165 165 165	ISOLATE  HK5  HK5  HK3  HK3  HK4  SA10  SN10  DK1  US6  IND3  P10  D1  US11  SN1  SN1  SN1  SN1  SN1  SN1  S	184 184 184 184 184 184 184 184 184 184	CLTIPASA

consensus

CLTiPaSA

#### FIGURE 7D

SEO ID NO:         ISOLATE           179         T9           178         US10           180         T2           177         T4           177-180         consensus	1 MSTNPKFORKTIRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRTTRKTSERSOPRGR 1 MSTNPKFÖRRTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPPLGVFATRKTSERSOPRGR 1 MSTIFFORFKTKRNTNRRPQDVKFPGGGGIVGGVYLLPRRGPRLGVFATRKTSERSOPRGR 1 MSTnPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR MSTnPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR
SEO ID NO:   ISOLATE   179	62 ROPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPEDPRHRSRNYGKV 62 ROPIPKDRRDTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPEDPRHRSRNYGKV 62 ROPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNYGKV 62 ROPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNYGKV ROPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNYGKV
SEO ID NO:   ISOLATE   179   T9	123 IDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 123 IDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 124 IDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 125 IDTLTCG1ADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS IDTLTCGfADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
SEO ID NO:   ISOLATE   179   T9   T9   178   US10   T2   177   T4   177-180   Consensus	184 CITTPASA 184 CITTPVSA 184 CITTPVSA CITTPVSA

# FIGURE 7E

SEO ID NO:         ISOLATE           183         DK11           184         SW3           181         T8           182         US1           185         DK8           181-185         consensus	1 MSTNPKPORKTKRNTNRRPODVKFFGGGOIVGGVYLLPRRGPRLGVRLTRKTSERSOPRGR 1 MSTNPKPORKTKRNTNRRPÖDVKFFGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR 1 MSTNPKPORKTKRNTNRRPÖDVKFFGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR 2 MSTNPKPORKTKRNTNRRPÖDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR 1 MSTNPKPORKTKRNTNRRPÖDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKSERSOPRGR MSTNPKPORKTKRNTNRRPÖDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKSERSOPRGR
SEO ID NO:   ISOLATE   183   DK11   184   SV3   181   TS1   185   US1   185   DK8   181-185   Consensus	62 ROPIPKDRRSTGKPWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHKSRNLGKV 62 ROPIPKDRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHKSRNLGKV 62 ROPIPKDRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGFTDPRHKSRNLGKV 62 ROPIPKDRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGFTDPRHKSRNLGKV 63 ROPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGFTDPRHKSRNLGKV ROPIPKDRRSTGK&WGKPGYPWPLYGNEGCGWAGWLLSPRGSTPWGFTDPRHKSRNLGKV
SEC ID NO:   ISOLATE   183   DR1   184   SW3   181   T8   182   US1   185   DR8   181-185   Consensus	123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 124 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 125 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
SEC ID NO: ISOLATE   183	184 CCTVPVSA 184 CFTVPVSA 184 CATVPVSA 184 CATVPVSA C-TVPVSA

#### FIGURE 7F

SEO ID NO 183 184 181 182 185 186 178 180 179 177	DK11 SW3 T8 US1 DK8 S83 US10 T2 T9 T4	1 MSTNPKPORKTKRNTNRRPODVKFPGGGOIVGGVYLLPRRGPRLGVRTTRKTSERSOPBGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGIVGGVYLLPRRGPRLGVRTRKTSERSOPBGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGIVGGVYLLPRRGPRLGVRATRKTSERSÖPPGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGIVGGVYLLPRRGPRLGVRATRKTSERSÖPPGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGIVGGVYLLPRRGPRLGVRATRKSESERSÖPPGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGIVGGVYLLPRRGPRLGVRATRKSESESÖPPGR 1 MSTNPKPÖRKTKRNTNRRPÖDVKFPGGGIVGGVYLLPRRGPRLGVRATRKTSERSÖPPGR
SEO ID NO 183 184 181 182 185 186 178 180 179 177	: ISOLATE     DK11     SW3     T8     US1     DK8     S83     US10     T2     T9     T4  Consensus	62 RQPIPKDRRSTGKDWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHKSRNLGKV 62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGFTDPRHRSRNLGKV 62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGGTDPRHRSRNLGKV 62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGFTDPRHRSRNLGKV 62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGFTDPRHRSRNLGKV 62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGFTDPRHRSRNLGKV 62 RQPIPKDRRTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGFTDPRHRSRNVGKV 63 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGFTDPHRSRNVGKV 64 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGFDPHPHRSRNVGKV 65 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGFDPHPHRSRNVGKV 66 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGFDPHPHRSRNVGKV 67 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGFDPHPHRSRNVGKV 68 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSFPSWGFDPHPHRSRNVGKV
SEO ID NO 183 184 181 182 185 186 178 179 177 177-186		123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 124 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 125 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 126 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS 127 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 128 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 129 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 120 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 121 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 123 IDTITCGFADLMGYPVVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 124 IDTITCGFADLMGYPVVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 125 IDTITCGFADLMGYPVVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 126 IDTITCGFADLMGYPVVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 127 IDTITCGFADLMGYPVVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 128 IDTITCGFADLMGYPVVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 129 IDTITCGFADLMGYPVVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS 129 IDTITCGFADLMGYPVVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
SEO ID NO 183 184 181 182 185 186 178 180 179 177	: ISOLATE DX11 SW3 T8 US1 DX8 S83 US10 T2 T9 T4	184 CCTVPVSA 184 CFTVPVSA 184 CFTVPVSA 184 CCTVPVSA 184 CCTVPVSA 184 CCTVPVSA 184 CTTPVSA

CitvPvSA

177-186

consensus

# FIGURE 7G

SEO ID NO:         ISOLATE           189         S2           187         HK10           190         DK12           188         S52	1 MSTLPKFÖRKTKRNTIRRPODIKFFGGGÖIVGGVYVLPRRGPRLGVRATRKTSERSÖPRGR 1 MSTLPKFÖRKTKRNTIRRPÖDVKFPGGGÖIVGGVYVLPRGPRLGVVATRKTSERSÖPRGR 1 MSTLPKFÖRKTKRNTIRRPÖDVKFPGGGÖIVGGVYVLPRGPRLGVVATRKTSERSÖPRGR 1 MSTLPKFÖRKTKRNTIRRPÖDVKFPGGGÖIVGGVYVLPRRGPRLGVRATRKTSERSÖPRGR
187-190 consensus	${ t MSTLPKPQRKTKRNTIRRPQDvKFPGGGQIVGGVYVLPRRGPRLGVRATRKTSERSQPRGR$
SEO ID NO:         ISOLATE           189         \$2           187         HX10           190         DX12           188         \$52           187-190         consensus	62 ROPIPKARRSEGRSWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV 62 ROPIPKARRSEGRSWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV 62 ROPIPKARRSEGRSWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV 62 ROPIPKARRSEGRSWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV ROPIPKARRSEGRSWAOPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
SEO ID NO:         ISOLATE           189         \$2           187         HK10           190         DK12           188         \$52           187-190         consensus	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS 123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS 123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS 123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVVALEDGINFATGNLPGCSFSIFLLALFS IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
SEO ID NO:   ISOLATE   189   S2   187   HK10   DK12   188   S52   187-190   Consensus	184 CLIHPAAS 184 CLIHPAAS 184 CLIHPAAS 184 CLIHPAAS CLIHPAAS

# FIGURE 7H

SEO ID NO: ISOLATE   194   195   28   195   26   196   27   191   24   197   24   197   24   197   24   197   25   191	1 MSTNPKPORKTKENTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRA-RKTSERSOPBGR 1 MSTNPKPORKTKENTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRABRKTSERSOPBGR 1 MSTNPKPORKTKENTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPBGR 1 MSTNPKPORKTKENTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPBGR 1 MSTNPKPORKTKENTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPBGR 1 MSTNPKPORKTKENTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPBGR 1 MSTNPKPORKTKENTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPBGR 1 MSTNPKPORKTKENTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR MSTNPKPQRKTKRNTNRRPMDVKFPGGGOIVGGVYLLPRRGPRLGVRATRKTSERSOPRGR
SEO ID NO: ISOLATE   194   25   25   27   27   27   27   27   27	62 ROPIPGARRSEGRSWAQPGYPWPLYGNEGGGMAGWLLSPEGSRPSWGGNDPRRRSRNLGKV 62 ROPIPKARRSEGRSWAQPGYPWPLYGNEGGGMAGWLLSPEGSRPSWGGNDPRRRSRNLGKV 62 ROPIPKARRSEGRSWAQPGYPWPLYGNEGGGMAGWLLSPEGSRPSWGPNDRRSRNLGKV 62 ROPIPKARRSEGRSWAQPGYPWPLYGNEGGGMAGWLLSPEGSRBSWGPNDRRRSRNLGKV 62 ROPIPKARRSEGRSWAQPGYPWPLYGNEGGGMAGWLLSPEGSRBSWGPNDPRRSRNLGKV 63 ROPIPKARDSEGRSWAQPGYPWPLYGNEGGGMAGWLLSPEGSRBWGMPDRRRSRNLGKV 64 ROPIPKARDSEGRSWAQPGYPWPLYGNEGGGMAGWLLSPEGSRBWGMNDRRRSRNLGKV 65 ROPIPKARDSEGRSWAQPGYPWPLYGNEGGGMAGWLLSPEGSRBWGMNDRRRSRNLGKV
SEO ID NO: ISOLATE   194   25   193   21   195   26   195   27   191   24   197   DK13   191   197   Consensus	RQPIPKARIBEGRSWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGpNDPRRRSRNLGKV  123 IDTLTCGFADLMGYIPLVGAPYGGVARALAHGVRA1EDGINYATGNLPGCSFSIFLLALIS 124 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLALIS 125 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLALIS 126 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLALIS 127 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLALIS 128 IDTLTCGFADLMGYIPVGAPVGGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLALIS 129 IDTLTCGFADLMGYIPVGAPVGGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLALIS 120 IDTLTCGFADLMGYIPVGAPVGGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLALIS
SEO ID NO: ISOLATE 194 25 193 21 195 26 195 27 196 27 191 24 197 DK13	184 CLTYPASA CLTYPASA

#### FIGURE 71

SEO ID NO:         ISOLATE           205         SA11           202         SA3           198         SA4           199         SA5           203         SA7           201         SA3           204         SA6	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQTVGGVYLLPRRGPRLGVRATRKTSERSQPRGR 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQTVGGVYLLPRRGPRLGVRATRKTSERSQPRGR 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGTVGGVYLLPRRGPRLGVRATRKTSERSQPRGR 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGTVGGVYLLPRRGPRLGVRATRKTSERSQPRGR 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGTVGGVYLLPRRGPRLGVRATRKTSERSQPRGR 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGTVGGVYLLPRRGPKLGVRATRKTSERSQPRGR 1 MSTNPKPQRKTGNNTNRPQDVKFPGGGTVGGVYLLPRRGPKLGVRATRKTSERQPRGR 1 MSTNPKPQRKTGNNTNRPQDVKFPGGGTVGGVYLLPRRGPRMGVRATRKTSERQPRGR
198-205 consensus	${\tt MSTNPKPQRKTkRNTNrRPQDVKFPGGGQIVGGVYLLPRRGPR1GVRATRKTSERSQPRGR$
SEO ID NO:   ISOLATE   205   SA11   202   SA3   198   SA4   199   SA5   200   SA7   201   SA1   SA1   204   SA6   SA6	62 ROPI PKAROPTGRSWGOPGYPHPF YANEGLGMAGHLLS PRGSRPHWGPND PRRISRNLGKV 62 ROPI PKAROPTGRSWGOPGYPHPLYANEGLGMAGHLLS PRGSRP BWGPND PRRISRNLGKV 62 ROPI PKAROPTGRSWGOPGYPHPLYANEGGMAGHLLS PRGSRPHWGPND PRRISRNLGKV 62 ROPI PKAROPTGRSWGOPGYPHPLYANEGGMAGHLS PRGSRPHWGFND PRRISRNLGKV 63 ROPI PKAROPTGRSWGOPGYPHPLYANEGGMAGHLS PRGSRPHWGFND PRRISRNLGKV 64 ROPI PKAGPGMAGHLSPHOFGRSPHWGFND PRRISRNLGKV
198-205 consensus	RQPIPKARQptGRSWGQPGYPWP1YANEGLgWAGWLLSPRGSRPnWGPNDPRRkSRNLGKV
SEO ID NO:         ISOLATE           2055         SA11           202         SA3           198         SA4           200         SA5           200         SA7           203         SA13           201         SA6           204         SA6	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRALEDGVNYATGNLPGCSFSIFILALLS 123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS 124 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS 125 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS 126 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS 127 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS 128 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS 129 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS 120 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
198-205 consensus	IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRvLEDGVNYATGNLPGCSFSIFiLALLS
SEO ID NO:   ISOLATE	184 CITVPALA 184 CITVPASA 184 CITPASA

198-205

consensus

CLtvPasA

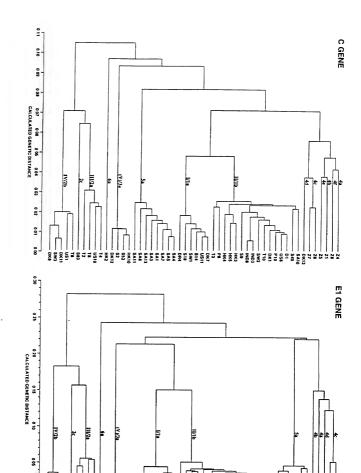
# PIGURE 73

LE NSTNPKPQRKTKRWTh:RPqDvKFPGGGQIVGGVYILPRRGPRIGVRatRK:SERSQPRGRRQPIPkaRtpeGt#WaqPGyPWPlYgnEGcgWAGW	NSTNEKPQRATKRNTNERPQDVKFPGGGQ I VGGVYLLPRRGFBLGVPLA EKTSERSQPRGRRQP I PKALF PBGR&NAQPQ-PHPLYGIEG - GHAGH NSTNEKPQRATKRNTNRRFDVFPGGGQ I VGGVYLLPRRGFBLGVPLAYRATRKTSERSQPBRRGPF I PKDARFGTGA MARF PROFPPLYGREGI GHAGH NSTLEKPQRATKRNTNRRFDVFPGGGQ I VGGYYLLPRRGPPLGVPLAYRYSERSQPBRRGPF I PKALF SEBSTANGPPPLYGREGGRAAN NSTREKPQRATKRATNRRFDVFPGGGQ I VGGYYLLLPRRGPPLGVPAR EKTSERSQPBRRGPF I PKALF SEBSTANGPPPLYGREGGRAAN NSTREKPQPRATKRATHREPGDVKFPGGGQ I VGGYYLLLPRRGPRATKTSERSQPBRRGP I PKALF SEBSTANGPFPPF I VANGGLANAGH NSTREKPQRATKRATHREPGDVKFPGGGQ I VGGYYLLLPRRGPRATKTSERSQPBRRGP I PKARQPCGSFBTPF I YANGGLANAGH NSTLEKRQCRATKRATHREPGDVKFPGGGQ I VGGYYLLLPRRGPRATRKTSERSQPBRGRRQF I PKARQPCGSFBTF I YANGGLANAGH NSTLEKRQCRATKRATHREPTDVKFPGGGQ I VGGYYLLLPRRGPRATRKTSERSQPBRGRRQF I PKARQPCGSFBTF I YANGGLANAGH	11. LISPEGSFB#WGPCDPREFSRNIGKVIDTITGGFADLMGY.PIUGGAPLGGVARALAHGWRVIEDGWYATGNIPGGGFSIFILALISCI tvPasa LISPEGSFB#WGPCDPRESRNIGKVIDTITGGFADLMGY.PUVGaPLGGARALAHGWRVIEDGWYATGNIPGGGFSIFILALISCI.TvPasa LISPEGSFB#WGPCDPRESRNIGKVIDTITGFADLMGY!PUVGaPLGGGNALAHGWRVIEDGJWYATGNIPGGGFSIFILALISCI.TvPvSA LISPEGSFB#WGPWDPRESRNIGKVIDTITGFADLMGY!PUVGAPWGGVNALAHGWAVEDGJWYATGNIAGGSFSIFILALISCI.TvPvSA LISPEGSFB#WGPWDPRESRNIGKVIDTITGFADLMGYIPVUGAPWGGVNALAHGWAVEDGJWYATGNIAGGSFSIFILALISCI.TVPASA LISPEGSFB#WGPWDPRESRNIGKVIDTITGFADLMGYIPVUGAPWGGVNALAHGWAVEDGJWYATGNIAGGSFSIFILALISCI.TVPASA LISPEGSFB#WGPWDPRESRNIGKVIDTITGFADLMGYIPVUGAPWGGVNALAHGWAVEDGJWYATGNIAGGSFSIFILALISCI.TVPASA LISPEGSFB#WGFWDPRESRNIGKVIDTITGFADLMGYIPVUGAPWGGVNALAHGWAVEDGJWYATGNIAGGSFSIFILALISCI.TVPASA LISPEGSFB#WGFWDPRESRNIGKVIDTITGFADLMGYIPVUGAPWGGVNALAHGWAVEDGJWYATGNIAGGSFSIFILALISCI.TVPASA LISPEGSFB#WGFWDPRESRNIGKVIDTITGFADLMGYIPVUGAPWGGVNALAHGWAVEDGJWYATGNIAGGSFSIFILALISCI.TVPRASA	
Genotype cons.	type 1 type 2 type 3 type 5 type 5	Genotype cons.  type 1 type 2 type 3 type 4 type 5 type 5 type 5 type 6 type 6 type 6 type 6	
SEQ ID NO: 155-206	155-176 177-186 187-190 191-197 198-205	SEQ ID NO. 155-206 155-176 177-186 191-197 198-206	

# FIGURE 7

06	NST-PKPORKTRRNTN-RP-QNKFPGGGGTVGEVTILPRRGPRTGVRSTGRRSGPRGRROPTR-ARF-PGG-R-ANG-PG-VP-UP T-G-G-G-G-MAGN			SE-RS-AQYL-GWCG SE-RS-AQYL-GWCG SE-RS-AQYL-GWCG DDT-RS-GQYL-GWLG	-GNCG	LLSPrGSFP8WGptDPRrrsRN1GKVIDT1TC9fADLMCYIPIYGaPLGGVARALAHGVRV1EDGVMYATGNIPGGSFSIFILALISCLtvPasa	11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	L1SV-VSA FLIH-AAS		## 5 - 00 ## 1   1   1   1   1   1   1   1   1
8e <sup>-</sup>	kaRrpeGrsWaqPGyPWP1YgnEg KA-RPE-RT-AQYL-GW		2222	2000	1Pa-RH-Aa·-YL-GNC 170, 180,	SFSIFILAL	× * × ×			
2-	KA-RPE-RT	00-831-KS	KA-RSE-RS KA-RSE-RS KA-RSE-RS	KA-RSE-RS KA-RSE-RS KA-RSE-RS	KA-aba-RH 170	NYATGNIPGO			,	
9-	OPRGRROPIF				169	HGVRVLEDGV	7777	743	A	
8-	ISTÆKPORKIRKNITNIRP-ØJ-VKFPGGGGIVGGVTLPRRGPRIGVRAKK LSENSOPRGRROP IP	¥ 0 0	 		100 110 120 130, 140 150, 160	PIGGVARALA	× 4 4 5 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7		× × × × × × × × × × × × × × × × × × ×	
0,	LPRRGPR1GVI	<u>.</u>	<u> </u>		92	MGY IP LYGAR		>	÷ • • • • • • • • • • • • • • • • • • •	* ÷ • * * * * * * * * * * * * * * * * *
9 —	GalvGGvYII	٠٠٠; 	÷÷÷÷		5	DTITC9fADL	25.22			****
<sub>8</sub> -	PobvKFPGG	>>>	>>>> •===	>>>>	120	-rsrnickvi		*****		****
ē-	KKTKRNTOFF -K-KNR-	****	****	****	7 13 E-	PSWGptDPR	S - PT - S	S-PI-	NA-S	S S C E
	MSTrPKPQ	c==		****	5	LLSPrGS	* - * * * * * * * * * * * * * * * * * *	~~~	~~~	****
	Genotype 1/1a 11/1b	111/2 <b>6</b> 1V/2b 2c	() ()	2258		Genotype	1/18 11/16 11/28	2c (V)/3s	222	35.88
	öl			197 194 198-205		SEG ID NO: (	155-160 161-176 177-180 181-185			





3 H H K K

US 10



Docket No. 2026-4116

#### IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s) : Jens Bukh, et al

Serial No. : 08/290,665

Group Art Unit: To be assigned

Filed

: August 15, 1994

Examiner: To be assigned

For

: NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE E ENVELOPE 1 AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN

DIAGNOSTIC METHODS AND VACCINES

Hon. Commissioner of Patents and Trademarks Washington, D.C. 20231

#### ASSOCIATE POWER OF ATTORNEY

Dear Sir:

Pursuant to the provisions of 37 CFR 1.33 and 1.34 and MPEP 402.02, the undersigned attorney of record hereby appoints the following as associate attorneys to prosecute this application, to receive the patent, and to transact all business in the Patent and

Trademark Office in connection with the above-identified application:

Kurt E. Richter (Reg. No. 24,052); Eugene Moroz (Reg. No. 25,237); William S. Feiler (Reg. No. 26,728); Israel Blum (Reg. No. 26,710); Bartholomew Verdirame (Reg. No. 28,483); Maria C. H. Lin (Reg. No. 29,323); Christopher E. Chalsen (Reg. No. 30,936); Eugene C. Raucidlo (Reg. No. 31,900); Mary J. Morry (Reg. No. 34,398); Michael M. Murray (Reg. No. 32, 537); Jean E. Shimotake (Reg. No. 36,273); Kathryn M. Brown (Reg. No. 34,556); Leslie A. Serunian (Reg. No. 35,353); Dorothy R. Auth (Reg. No. 36,454); Richard W. Bork (Reg. No. 36,459); M. Caragh Noone (Reg. No. 37,7341) of Morgan & Finnegan whose address is: 345 Park Avenue, New York, New York 10154.

Respectfully submitted.

Date	12/15/94	
		Ann S. Hobbs

Registration No. 36,830

Patent Branch Office of Technology Transfer National Institutes of Health

Box 13

6011 Executive Boulevard, Suite 325 Rockville, MD 20852

Gota

Tel. No. (301) 496-7056

Docket 2026-4116

#### COMBINED DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name, the information given herein is true, that I believe I am the original, first and sole (if only one name is listed below) or an or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE ENVELOPE I AND CORE GENES OF ISOLATES OF HEPATITIS C VIRUS AND THE USE OF REAGENTS DERIVED FROM THESE SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES.

	A STATE OF THE STA
which is described in: [ ] the attached application or	[ ] PCT International Application No. filed [X] the specification in application Serial No. 08/290,665 filed August 15, 1994 (if applicable) and amended on
	(it applicable) and amended on

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information known to me which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56 (a).

I hereby claim foreign priority benefits under Title 35 United States Code, § 119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign applications(s) for patent or inventor's certificate or any PCT international applications(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed.

COUNTRY	APPLICATION	DATE OF FILING (day, month, year)		Y CLAIMED IS USC § 119
			[] Yes	[] No
			[] Yes	[] No
			[] Yes	[] No

I hereby claim the benefit under Title 35, United States Code \$120 of any United States application(s) or PCT International application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, \$112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, \$1.56(a) which occurred between the filing date of the prior application(s) and the national or PCT international filing date of this application.

Application Serial No.	Filing Date	Status: patented, pending, abandoned
08/086,428	29 June 1993	pending

I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

James C. Haight, Reg. No. 25,588; Gloria Richmond, Reg. No. 30,416; Robert Benson, Reg. No. 33,612; Jack Spiegel, Reg. No. 34,477; Laurence J. Hyman, Reg. No. 35,551; Denise C. Bernstein, Reg. No. 35,787; Susan S. Rucker, Reg. No. 35,762; David R. Sadowski, Reg. No. 32,808 and Ann S. Hobbs, Reg. No. 36,830 and Arthur J. Cohn, Reg. No. 37,800 all of the Office of Technology Transfer, National Institutes of Health, 6011 Executive Boulevard, Suite 325, Rockville, MD 20852

I further direct that all correspondence concerning this application be directed to:

Patent Branch
Office of Technology Transfer
National Institutes of Health
Box 13
6011 Executive Boulevard, Suite 325
Rockville, MD 20852
Telephone: (301) 496-7056

Docket No. 2026-4116

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full Name of first joint inventor: Jens Rukh

2010 2010	
Inventor's signature: 100 / mulch	Date: 12/15/94
Country of Citizenship: Denmark	
Residence: 5805 Sonoma Road, Bethesda, Maryland 20817, U.S.A.	
Post Office Address: 5805 Sonoma Road, Bethesda, Maryland 20817, U.S.A.	
Full Name of second joint inventor: <u>Roger H. Miller</u>	
Inventor's signature: Permitty - Mill	Date: 13-115/90,
Country of Citizenship: United States of America	
Residence: 15504 White Willow Lane, Rockville, Maryland 20853, U.S.A.	
à.	
Post Office Address: 15504 White Willow Lane, Rockville, Maryland 20853, U.S.A.	
Full Name of third joint inventor: Robert H. Purcell	
Inventor's signature: Keled Www.	Date: 12/15/74
Country of Citizenship: United States of America	
Residence: 17517 White Grounds Road, Boyds, Maryland 20841, U.S.A.	
Post Office Address: 17517 White Grounds Road, Boyds, Maryland 20841, U.S.A.	